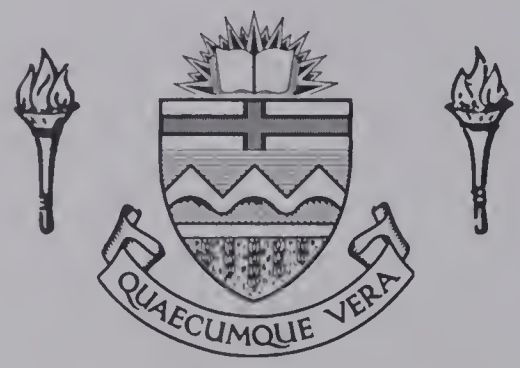


The University of Alberta
Printing Department
Edmonton, Alberta

THE UNIVERSITY OF ALBERTA

CONTRIBUTIONS TO THE AMINO ACID SEQUENCE OF α-LYTIC PROTEASE

by

MICHAEL JIMMY RICHARD DZWINIEL

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF SCIENCE

DEPARTMENT OF BIOCHEMISTRY

EDMONTON, ALBERTA

JULY, 1968

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "CONTRIBUTIONS TO THE AMINO ACID SEQUENCE OF α-LYTIC PROTEASE", submitted by Michael Jimmy Richard Dzwiniel in partial fulfilment of the requirements for the degree of Master of Science.

ABSTRACT

The isolation and characterization of the α-lytic protease of *Sorangium sp.* by Whitaker and his collaborators demonstrated that the "active serine" sequence of this enzyme was homologous with the mammalian serine proteases and different from the corresponding bacterial proteases of *Bacillus subtilis* and *Aspergillus oryzae*. Further work by Smillie, Whitaker and Kaplan showed that this homology was also present in the sequence about the single histidine residue of the enzyme and that the catalytic activity had several properties in common with chymotrypsin.

To elucidate the extent of homology between this enzyme and the mammalian proteases, the determination of its complete amino acid sequence has been undertaken in this laboratory. Towards this end, peptides from tryptic digests of the reduced and S-carboxymethylated enzyme and from chymotryptic digests of the reduced and S-aminoethylated enzyme have been purified, characterized and, in many cases, sequenced. These results, together with the observations of Drs. Nagabhushan and Olson of this laboratory on the peptides arising from the tryptic digestion of the reduced and S-aminoethylated enzyme, and those of Dr. Whitaker of Ottawa on the fragments arising from the cyanogen bromide cleavage of the enzyme, have permitted the tentative assignment of all amino acid residues of the protein into six sequences. The appropriate overlapping peptides for these six fragments have not yet been isolated. The largest of the sequences (133 residues) includes both the "active serine" sequence and the C-terminus of the molecule.

When this sequence is aligned with the sequences of chymotrypsinogens A and B and trypsinogen in such a manner as to maximize homologies about the "active serines", the half-cystines and the C-termini, it is found that the common pattern of invariant non-polar residues present in the three mammalian proteases is, to a considerable extent, also present in α-lytic protease. This is true even though identity of sequence is restricted to the region about the "active serine". This indication of a common tertiary structure in at least a portion of the α-lytic protease molecule and the mammalian proteases is further substantiated by the existence of a disulfide bridge in a similar position in all four enzymes. These homologies suggest that α-lytic protease and the mammalian enzymes are descended from a common ancestral protein.

ACKNOWLEDGEMENTS

I would sincerely like to express my gratitude to my supervisor, Dr. L.B. Smillie, for his interest and guidance during these studies.

I wish to thank Mr. E. Paradowski for the numerous amino acid analyses which he performed.

I am indebted to Dr. N. Nagabhushan and Dr. M. Olson for providing their unpublished data for use in this thesis. The results of cyanogen bromide cleavage from Dr. D.R. Whitaker are gratefully acknowledged.

I am most appreciative of a pleasant association with Dr. J.S. Colter and the staff of the Department of Biochemistry. Thanks is extended to Miss V. Lindgren who typed the thesis.

TABLE OF CONTENTS

LIST OF TABLES AND ILLUSTRATIONS

LIST OF ABBREVIATIONS

DFP	diisopropylfluorophosphate
tris	tris (hydroxymethyl) aminomethane
TPCK	L-1-tosylamido-2-phenylethyl chloromethyl ketone
TLCK	1-chloro-3-tosylamido-7-amino-2-heptanone
Kv	kilovolt
Dansyl chloride	1-dimethylaminonaphthalene-5-sulfonyl chloride
ammediol	2-amino-2-methyl-1,3-propanediol
EDTA	ethylenediaminetetraacetate
sp.	species
ATEE-ase	N-acetyl-L-tyrosine ethyl esterase

All amino acids are abbreviated by the first three letters except isoleucine (Ile), glutamine (Gln), asparagine (Asn) and tryptophan (Trp). Glx and Asx stand for either the free acid or amide form of glutamic acid and aspartic acid respectively.

CHAPTER I

INTRODUCTION

The most thoroughly studied group of enzymes throughout the years has been the proteolytic enzymes, which are found almost universally in nature. Not only are they of interest in themselves, but due to their specific action in hydrolysis they have been used extensively for degrading other proteins and peptides. Of great importance within this group of enzymes is the class of "serine" proteases, so called because they all have an "active" serine residue in the sequence of amino acid residues around the active site. Since this group of enzymes is in itself important and since it bears directly on the subject of this thesis, a brief review of the relevant knowledge of the structures and activities of the proteolytic enzymes with an emphasis on the serine proteases is in order.

Classification

The proteolytic enzymes are found universally in the plant, animal and microbial worlds. The better characterized enzymes have been categorized on the basis of their specificity in hydrolysis (3). However, a systematic nomenclature covering all peptide hydrolases is not possible at present, owing to their overlapping specificities. The separate identity of some of them seems to be somewhat doubtful. According to the recommendations of the Report of the Commission on Enzymes of the International Union of Biochemistry (3), those enzymes

acting on peptide bonds, peptide hydrolases, (serial number 3.4) have been sub-classified into E. C. 3.4.1 α-aminopeptide aminoacidhydrolases (or the aminopeptidases), E. C. 3.4.2 α-carboxypeptide aminoacidhydrolases (the carboxypeptidases), E. C. 3.4.3 dipeptide hydrolases (or the dipeptidases) and E. C. 3.4.4 peptide peptidohydrolases (the endopeptidases which include trypsin, pepsin, chymotrypsin, etc.)

The endopeptidases can be more specifically categorized on the basis of mechanism of catalysis to form a number of alternate, and convenient groupings (48):

(a) the "serine" proteases in which a serine residue is acylated during the hydrolysis.

(b) the "acid" proteases or pepsins, characterized by their very low pH optima. Hydrolysis in these enzymes is believed to be the result of simultaneous attack by two enzyme carboxyl groups on both the amine and carbonyl moieties of the substrate peptide bond.

(c) the "thiol" proteases or plant proteases of which papain is an example. Here a thiol group is acylated during catalysis.

(d) the "intracellular" proteases or cathepsins. An active thiol group is suggested in some of these enzymes but no definite classification of this type should be made until the problem of impure preparations has been dealt with. The term "cathepsin" seems to imply a homogeneity of properties or functions which is unwarranted. Hartley (4) suggests that a number of these enzymes may later be classified in another category.

(e) the "metal" proteases in which a metal ion is contained within the enzyme molecule and is essential for activity.

Characterization of these classes is a relatively easy task due to their distinctive properties. The serine proteases are readily inactivated by organophosphate inhibitors (for example diisopropylfluorophosphate , or DFP), the thiol proteases by thiol reactive reagents (like p-hydroxymercuribenzoate, or p-HMB), the pepsins by deviation from low pH, and the metal proteases by complexing the metal ion with agents such as ethylenediaminetetraacetate (EDTA). Characterization of the cathepsins is more difficult since no basic property of this class exists.

A partial list of serine proteases has been compiled (Table 1-1) but since not all known endopeptidases can be classified satisfactorily, many more enzymes may belong to this class. It has been suggested that agavain may not be a true serine protease (5).

The mechanism of hydrolysis of the serine proteases

The presently favored mechanism is basically that presented by Cunningham (6) for chymotrypsin catalyzed reactions. It seems to be valid for the tryptic reactions and probably for the other serine proteases as well. An extensive compilation of the evidence for this mechanism has been made (7-13). The overall reaction is a three step process:

(a) the formation of the enzyme-substrate complex (ES)

(b) acylation of the active serine hydroxyl group by the carbonyl moiety of the substrate forming the acyl-enzyme

Table 1-1

Serine Proteases

Enzyme	Source	Molecular Wt.	Active Residue
Trypsin	bovine	23,700	Ser
	canine		
	equine		
	rat anionic		
	turkey	22,500	
	salmon		
Chymotrypsin A	bovine	25,767	Ser
	porcine	22,700	
	dogfish	24,500	
Chymotrypsin B	bovine anionic	26,000	Ser
	porcine		
Chymotrypsin C (fr. II of procarboxypeptidase A complex	bovine	25,000	
Chymotrypsin C	porcine	31,800	
Chymotrypsin	canine		
	chicken	20,000-26,000	
Elastase	porcine	25,000	Ser
Thrombin	bovine	15,000-20,000	Ser
Plasmin or Fibrinolysin	bovine		Ser
Subtilisin Novo	*B. subtilis*	27,600	Ser
Carlsberg	*B. subtilis*	27,600	Ser
BPN'(Nagarse)	*B. subtilis*	27,600	Ser
Aspergillopeptidase A	*Aspergillus oryzae*		Ser
α-Lytic Protease	*Sorangium sp.*	20,100	Ser

Other Possible Serine Proteases

Cocoonase (trypsin-like)	silk worm	25,000
Intracellular protease	*Streptomyces moderatus*	20,000
Renin	bovine kidney	
ATEE-ase protease	*B. licheniformis*	
Trypsin-like protease	sea urchin egg	
Enterokinase	intestinal mucosa	
Agavain	sisal extract	56,000 (1 atom Fe^{++})

intermediate (ES') and P_1 the alcohol portion of the substrate. This process was first suggested in 1950 for the acetylcholinesterase catalyzed hydrolyses (14) and later applied to the chymotrypsins (15).

(c) deacylation and release of the carboxylic acid portion of the substrate P_2.

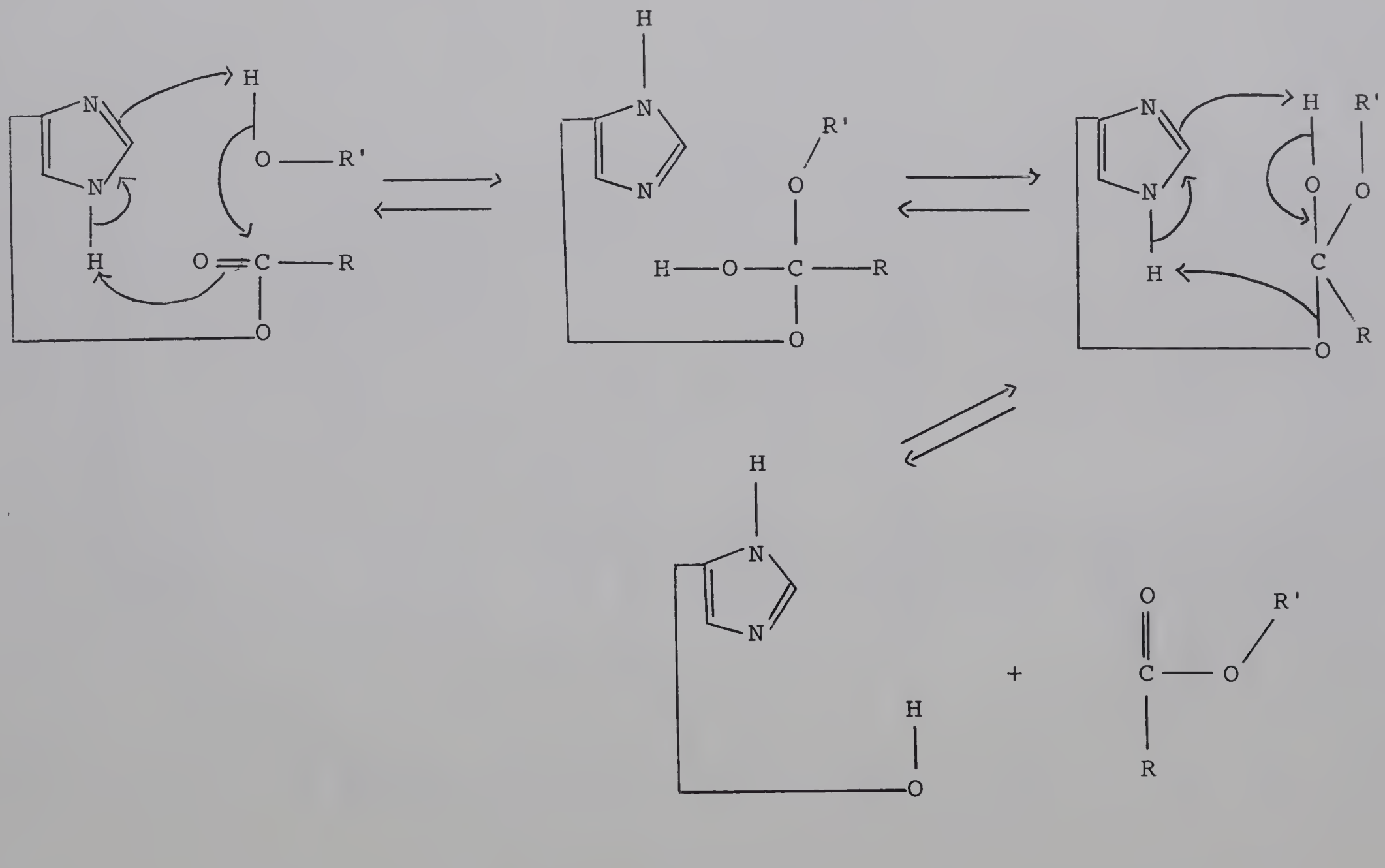
$$E + S \rightleftharpoons ES \longrightarrow ES' \longrightarrow E + P_2$$
$$+P_1$$

The exact molecular mechanism is still in doubt. A great many variations of mechanisms involving one serine and one or two histidines are present in the literature. The most complete study is that reported by Bender and Kézdy (16) who propose the mechanism shown in Figure 1-1 for the deacylation of α-chymotrypsin. The deacylation step is shown since a large amount of mechanistic information is available for this step and acylation is simply the microscopic reverse of it.

The mechanism agrees with all known experimental data pertinent to chymotrypsin catalyses including: (i) pH dependencies; (ii) the acyl-enzyme is a serine ester; (iii) an imidazole is involved in acylation and deacylation: (iv) the acylation and deacylation are nucleophilic reactions; (v) no detectable intermediate is formed in deacylation since tetrahedral addition compounds are unstable; (vi) the requirement of microscopic reversibility is met; (vii) the mechanism is simple, straightforward, and utilizes the unique ability of imidazole to serve simultaneously as a general base and a general acid; (viii) the reaction has the attributes of a

Figure 1-1

The Mechanism for α-Chymotrypsin Hydrolyses



concerted reaction which should enhance its kinetic efficacy; (ix) all transition states should be neutral, predicting no effect of ionic strength or dielectric constants on the rates, as found experimentally; and (x) due to the contribution from general acid catalysis, the enzymatic deacylation should be faster than a corresponding intramolecular general basic catalysis, as found experimentally.

The negative aspect of this mechanism is that the steric requirements of the reaction between the imidazole molecule, the substrate, and the serine hydroxyl group are not met. Bender has also proposed a mechanism involving two histidines (13) but the X-ray crystallographic data of Blow (17) show the second histidine of chymotrypsin removed from the active centre. This evidence along with the discovery of the bacterial serine protease Sorangium sp. α-lytic protease, which apparently utilizes the same catalytic mechanism and is functionally competent with only one histidine, have made this mechanism less popular.

The two histidine hypothesis was originally proposed only because of one of the homologous sequences in trypsin and chymotrypsin which contained two histidines, and not because of any experimental kinetic evidence for such a mechanism.

Zymogen activation

Precursor zymogens activated by specific proteolytic cleavages have been discovered for all mammalian serine proteases but not for the bacterial proteases. The mechanisms of zymogen activation by trypsin or some proteolytic enzyme

with trypsin-like activity have many common features. For example the initial event in all the activations is a proteolytic cleavage of a peptide bond in the zymogen molecule. In both trypsin and chymotrypsin this initial cleavage results in the formation of a new N-terminal isoleucine residue. The ionized form of this amino acid is necessary for the activity of both trypsin and chymotrypsin. The data of Blow et al. (17) suggest that the N-terminal isoleucine forms a specific salt linkage with the carboxyl group of aspartic acid-197 adjacent to the active serine residue (see Table 1-2) thus stabilizing the conformation of the active centre. The bonds first split in trypsinogen and chymotrypsinogen (-Lys$\overset{\downarrow}{-}$Ile- and -Arg$\overset{\downarrow}{-}$Ile- respectively) are in exactly the same location in the two enzymes (residues 15-16 according to Hartley's homologous numbering system) (18).

These similarities between the activation of trypsinogen and chymotrypsinogen suggest that the mechanism may be common in many respects for all the mammalian serine proteases.

Comparison of total sequences

The amino acid sequences for chymotrypsinogen A (19), chymotrypsinogen B (20) and trypsinogen (21) are now known and these proteins show a large proportion of common amino acid sequences. Chymotrypsinogen A and trypsinogen when aligned give coincidences of amino acids corresponding to 40% of the total sequence. If similar amino acids (for example lysine and arginine, aspartic acid and glutamic acid, serine and threonine etc.) are equated, the homologous areas include 51% of the protein. From this large proportion of

homology in the primary sequence it can be concluded that the two enzymes had the same ancestral enzyme and each has retained a considerable portion of the original structure. The sequence of chymotrypsinogens A and B present even a greater proportion of coincident residues. An 80% homology here suggests that they also had a common ancestor but have deviated from it more recently than trypsinogen. Table 1-2 compares the structures of some of the serine proteases mentioned above. The sequences of trypsinogen, chymotrypsinogen A and chymotrypsinogen B were from the work of Smillie *et al*. (49) while the results for elastase were taken from an earlier article by Hartley and co-workers (18).

X-ray crystallographic studies on several proteins have shown that one of the most striking common structural similarities is the near total exclusion of polar residues from the interiors of the molecules. Of the 245 residues in each of chymotrypsinogens A and B and the 229 in trypsinogen, 100 residues are invariably non polar and may be tentatively identified as interior (49). Thus additional support is provided for the extensive similarities in the three dimensional structure of these proteins previously suggested by the homology of their disulfide bridges.

Areas of greatest homology

Certain specific areas of the serine proteases show a high proportion of homologous residues. These include the areas around the disulfide bonds, the active centre and the histidine residues. These are shown in Tables 1-3 and 1-4.

Table 1-2

Amino Acid Sequences of Porcine Elastase (E), Bovine Trypsinogen (T), Chymotrypsinogen A (CA) and Chymotrypsinogen B (CB)*

*Identical residues are underlined unless they are only common to CA and CB. Disulfide bridges are lettered A to G. Asx indicates aspartic acid or asparagine and glx stands for glutamic acid or glutamine. The "overlap" between residues 188 and 189 of elastase is uncertain.

	E														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
E:															
T:										Val	Asp	Asp	Asp	Asp	Lys
CA:	Cys	Gly	Val	Pro	Ala	Ile	Gln	Pro	Val	Leu	Ser	Gly	Leu	Ser	Arg
CB:	Cys	Gly	Val	Pro	Ala	Ile	Gln	Pro	Val	Leu	Ser	Gly	Leu	Ala	Arg

							G								
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
E:															
T:	Ile	Val	Gly	Gly	Tyr	Thr	Cys	Gly	Ala	Asn	Thr	Val	Pro	Tyr	Gln
CA:	Ile	Val	Asn	Gly	Glu	Glu	Ala	Val	Pro	Gly	Ser	Trp	Pro	Trp	Gln
CB:	Ile	Val	Asn	Gly	Glu	Asp	Ala	Val	Pro	Gly	Ser	Trp	Pro	Trp	Gln

												A			
	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
E:								Trp	Ala	His	Thr	Cys	Gly	Gly	Thr
T:	Val	Ser	Leu	Asn	-	-	Ser	Gly	Tyr	His	Phe	Cys	Gly	Gly	Ser
CA:	Val	Ser	Leu	Gln	Asp	Lys	Thr	Gly	Phe	His	Phe	Cys	Gly	Gly	Ser
CB:	Val	Ser	Leu	Gln	Asp	Ser	Thr	Gly	Phe	His	Phe	Cys	Gly	Gly	Ser

													A		
	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
E:	Leu								Thr	Ala	Ala	His	Cys	Val	Asp
T:	Leu	Ile	Asn	Ser	Gln	Trp	Val	Val	Ser	Ala	Ala	His	Cys	Tyr	Lys
CA:	Leu	Ile	Asn	Glu	Asn	Trp	Val	Val	Thr	Ala	Ala	His	Cys	Gly	Val
CB:	Leu	Ile	Ser	Glu	Asp	Trp	Val	Val	Thr	Ala	Ala	His	Cys	Gly	Val

	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
E:	Arg	Glx													
T:	Ser	Gly	Ile	Gln	Val	Arg	Leu	-	Gly	Gln	-	Asp	Asn	Ile	Asn
CA:	Thr	Thr	Ser	Asp	Val	Val	Val	Ala	Gly	Glu	Phe	Asp	Gln	Gly	Ser
CB:	Thr	Thr	Ser	Asp	Val	Val	Val	Ala	Gly	Glu	Phe	Asp	Gln	Gly	Leu

	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
E:															
T:	Val	Val	Glu	Gly	Asn	Gln	Gln	Phe	Ile	Ser	Ala	Ser	Lys	Ser	Ile
CA:	Ser	Ser	Glu	Lys	-	Ile	Gln	Lys	Leu	Lys	Ile	Ala	Lys	Val	Phe
CB:	Glu	Thr	Glu	Asp	-	Thr	Gln	Val	Leu	Lys	Ile	Gly	Lys	Val	Phe

	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105
E:							Ser	Asn	Thr	Leu					
T:	Val	His	Pro	Ser	Tyr	Asn	Ser	Asn	Thr	Leu	Asn	Asn	Asp	Ile	Met
CA:	Lys	Asn	Ser	Lys	Tyr	Asn	Ser	Leu	Thr	Ile	Asn	Asn	Asn	Ile	Thr
CB:	Lys	Asn	Pro	Lys	Phe	Ser	Ile	Leu	Thr	Val	Arg	Asn	Asp	Ile	Thr

Table 1-2 (continued)

	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
E:															
T:	Leu	Ile	Lys	Leu	Lys	Ser	Ala	Ala	Ser	Leu	Asn	Ser	Arg	Val	Ala
CA:	Leu	Leu	Lys	Leu	Ser	Thr	Ala	Ala	Ser	Phe	Ser	Gln	Thr	Val	Ser
CB:	Leu	Leu	Lys	Leu	Ala	Thr	Pro	Ala	Gln	Phe	Ser	Glu	Thr	Val	Ser

			E						F						
	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135
E:												Ala	Asn	Asn	Ser
T:	Ser	Ile	Ser	Leu	Pro	Thr	-	Ser	Cys	Ala	Ser	-	Ala	Gly	Thr
CA:	Ala	Val	Cys	Leu	Pro	Ser	Ala	Ser	Asp	Asp	Phe	Ala	Ala	Gly	Thr
CB:	Ala	Val	Cys	Leu	Pro	Ser	Ala	Asp	Glu	Asp	Phe	Pro	Ala	Gly	Met

		D													
	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150
E:	Pro	Cys	Tyr												
T:	Gln	Cys	Leu	Ile	Ser	Gly	Trp	Gly	Asn	Thr	Lys	Ser	Ser	Gly	Thr
CA:	Thr	Cys	Val	Thr	Thr	Gly	Trp	Gly	Leu	Thr	Arg	Tyr	Thr	Asn	Ala
CB:	Leu	Cys	Ala	Thr	Thr	Gly	Trp	Gly	Lys	Thr	Lys	Tyr	Asn	Ala	Leu

								G							
	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165
E:															
T:	Ser	Tyr	Pro	Asp	Val	Leu	Lys	Cys	Leu	Lys	Ala	Pro	Ile	Leu	Ser
CA:	Asn	Thr	Pro	Asp	Arg	Leu	Gln	Gln	Ala	Ser	Leu	Pro	Leu	Leu	Ser
CB:	Lys	Thr	Pro	Asp	Lys	Leu	Gln	Gln	Ala	Thr	Leu	Pro	Ile	Val	Ser

				B											
	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180
E:		Ala	Ile	Cys	Ser	Ser	Ser	Ser	Ser	Tyr					
T:	Asn	Ser	Ser	Cys	Lys	Ser	Ala	Tyr	Pro	Gly	Gln	Ile	Thr	Ser	Asn
CA:	Asn	Thr	Asn	Cys	Lys	Lys	Tyr	Trp	Gly	Thr	Lys	Ile	Lys	Asp	Ala
CB:	Asn	Thr	Asp	Cys	Arg	Lys	Tyr	Trp	Gly	Ser	Arg	Val	Thr	Asp	Val

				B										C	
	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195
E:	Met	Val	Cys	Ala	Gly		Gly	Asp	Gly	Val	Arg	Ser	Gly	Cys	Gln
T:	Met	Phe	Cys	Ala	Gly	Tyr	Leu	Glu	Gly	Gly	Lys	Asn	Ser	Cys	Gln
CA:	Met	Ile	Cys	Ala	Gly	Ala	-	Ser	Gly	Val	-	Ser	Ser	Cys	Met
CB:	Met	Ile	Cys	Ala	Gly	Ala	-	Ser	Gly	Val	-	Ser	Ser	Cys	Met

									D						
	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210
E:	Gly	Asp	Ser	(Gly	Gly	Pro)	Leu	His	Cys	Leu	Val	Asn	Gln	Tyr	
T:	Gly	Asp	Ser	Gly	Gly	Pro	Val	Val	Cys	Ser	Gly	Lys	-	-	-
CA:	Gly	Asp	Ser	Gly	Gly	Pro	Leu	Val	Cys	Lys	Lys	Asn	Gly	Ala	Trp
CB:	Gly	Asp	Ser	Gly	Gly	Pro	Leu	Val	Cys	Gln	Lys	Asn	Gly	Ala	Trp

													C		
	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225
E:								Val	Ser	Arg	Leu	Gly	Cys	Asn	Val
T:	-	Leu	Gln	Gly	Ile	Val	Ser	Trp	Gly	Ser	-	Gly	Cys	Ala	Gln
CA:	Thr	Leu	Val	Gly	Ile	Val	Ser	Trp	Gly	Ser	Ser	Thr	Cys	Ser	Thr
CB:	Thr	Leu	Ala	Gly	Ile	Val	Ser	Trp	Gly	Ser	Ser	Thr	Cys	Ser	Thr

Table 1-2 (continued)

	226	227	228	229	230	231	232	233	234	235	F 236	237	238	239	240
E:	Thr	Arg	Lys	Pro	Thr	Val	Phe								
T:	Lys	Asn	Lys	Pro	Gly	Val	Tyr	Thr	Lys	Val	Cys	Asn	Tyr	Val	Ser
CA:	Ser	Thr	-	Pro	Gly	Val	Tyr	Ala	Arg	Val	Thr	Ala	Leu	Val	Asn
CB:	Ser	Thr	-	Pro	Ala	Val	Tyr	Ala	Arg	Val	Thr	Ala	Leu	Met	Pro

	241	242	243	244	245	246	247	248	249
E:									
T:	Trp	Ile	Lys	Gln	Thr	Ile	Ala	Ser	Asn
CA:	Trp	Val	Gln	Gln	Thr	Leu	Ala	Ala	Asn
CB:	Trp	Val	Gln	Glu	Thr	Leu	Ala	Ala	Asn

Table 1-3

Active Centre Sequences of Some Serine Proteases

Enzyme	Active Centre Sequence	Reference
Chymotrypsin A	- Gly Asp Ser* Gly -	23
Chymotrypsin B	- Gly Asp Ser* Gly -	20
Trypsin	- Gly Asp Ser* Gly -	25
Elastase	- Gly Asp Ser* Gly -	28
Thrombin	- Gly Asp Ser* Gly -	27
α-Lytic Protease	- Asp Ser* Gly -	29
Liver Ali-Esterase	- Gly Glu Ser* Ala -	30
Pseudocholinesterase	- Gly Glu Ser* Ala -	31
Subtilisin	- Thr Ser* Met Ala -	33
Aspergillus Protease	- Thr Ser* Met Ala -	32

Table 1-4

Histidine Disulfide Structures for Some Proteolytic Enzymes[a]

	39	40	41	42	43	44	45	46	47	48	49	50
Chymotrypsin A	Phe	His	Phe	Cys	Gly	Gly	Ser	Leu	Ile	Asn	Glu	Asn
Chymotrypsin B	Phe	His	Phe	Cys	Gly	Gly	Ser	Leu	Ile	Ser	Glu	Asp
Trypsin	Tyr	His	Phe	Cys	Gly	Gly	Ser	Leu	Ile	Asn	Ser	Gln
Elastase	Ala	His	Thr	Cys	Gly	Gly	Thr	Leu				
α-Lytic Protease				Cys	Ser	Val	Gly	Phe				

	51	52	53	54	55	56	57	58	59	60	61	62	63
Chymotrypsin A	Trp	Val	Val	Thr	Ala	Ala	His	Cys	Gly	Val	Thr	Thr	Ser
Chymotrypsin B	Trp	Val	Val	Thr	Ala	Ala	His	Cys	Gly	Val	Thr	Thr	Ser
Trypsin	Trp	Val	Val	Ser	Ala	Ala	His	Cys	Tyr	Lys	Ser	Gly	Ile
Elastase				Thr	Ala	Ala	His	Cys	Val	Asp	Arg	Glx	
α-Lytic Protease	Phe	Val	Thr	Ala	Gly	His	Cys	Gly	Thr	Val	Asn	Ala	

[a]The disulfide bridge is between residues corresponding to half-cystines 42 and 58 of chymotrypsin in each case.

(a) Common "active serine" sequences

The phosphorylating reaction using organophosphate inhibitors has provided a method of isolating active centre peptides. Isotopically labelled DFP (diisopropylfluorophosphate) and labelled Sarin (methylisopropylfluorophosphate) have been used. Most common is DF^{32}P first used for this purpose by Schaffer et al. (22). The sequence in chymotrypsin of -Asp Ser* Gly- was found by Turba and Gundlach (23). The "active serine" sequence of trypsin was correctly reported by Walsh and Neurath (25) to be -Gly Asp Ser* Gly Gly Pro-. The similarity of the sequence -Asp Ser* Gly- was quickly recognized and soon found to be common to other members of the class of serine proteases. Table 1-3 shows the "active serine" sequence for some serine proteases.

The "active sequence" of -Gly Asp Ser* Gly- in the serine proteases and the great similarity of the sequence -Gly Glu Ser Ala- in the aliphatic esterases led to speculation about the role of these particular sequences of amino acid residues in catalysis, but no experimental evidence has been deduced to support any such role for residues other than the serine. Studies of the bacterial proteases subtilisin and aspergillopeptidase, which have an "active serine" sequence of -Thr Ser* Met Ala- present evidence against any theory suggesting a necessity for the -Gly Asp Ser* Gly- sequence for activity. Only the serine seems to be essential. Other existing possibilities include a similar mechanism involving slightly different structures, and a different mechanism altogether. The role of these residues might not be directly involved

with catalysis. For example the X-ray crystallographic data of Blow (17) shows that the aspartic acid residue is important in the activation of chymotrypsinogen since the ionized carboxyl group forms a specific salt link with the N-terminal isoleucine residue, thus stabilizing the conformation of the active centre and assisting activation of the zymogen.

(b) The disulfide bridges

A common feature of the mammalian serine proteases is the high content of disulfide bonds. This tends to stabilize the conformation of the protein and is particularly important in the areas of the active site and substrate binding site. Therefore it is not surprising that these disulfide areas have similar amino acid sequences.

Recent work by Sigler and Blow (50) has shown that the two extra disulfide bridges of trypsin can be added to the chymotrypsin molecule without resulting in distortion of the chains. This suggests even more strongly that there is a considerable similarity in the three-dimensional structure of these two proteins.

(c) The histidine residues

Substrate analogue alkylating reagents have been used to show participation of histidine in the catalytic mechanism of the serine proteases. For example, Schoellmann and Shaw (34) demonstrated involvement of a histidine residue in the catalytic activity of α-chymotrypsin using l-tosylamido-2-phenylethyl chloromethyl ketone (TPCK). The phenylalaninyl side chain and the tosylamido group of this reagent enable it to be bound to the chymotrypsin molecule while the chloromethyl

ketone group forms a covalent bond with a residue in the active site, thus making it possible to isolate the histidine associated with the activity of the enzyme. The particular residue, which when reacted in this way rendered the enzyme inactive, was shown by Smillie and Hartley (43) to be histidine-57. Similar use of the chloromethyl ketone derivative of tosyl-L-lysine, 1-chloro-3-tosylamido-7-amino-2-heptanone (TLCK), led Mares-Guia, Shaw and Cohen to show histidine participation in the catalytic action of trypsin (35).

The most conclusive evidence for one histidine in the mechanism is the X-ray crystallographic data of Blow et al. (17) which show the histidine-57 of α-chymotrypsin "pointing towards" the active serine residue and approaching within 5 Å of it. The other histidine (histidine-40) appeared to be at least 13 Å away from the active site histidine and therefore is not likely to participate in the catalytic reaction.

Bacterial serine proteases

The active sequence of the subtilisins first showed that these bacterial enzymes were not homologous with the mammalian serine proteases (33). Further investigation showed that they also differed in another way; the bacterial enzymes contained no disulfide bonds. The two subtilisins were 70% homologous with each other and neither showed any homology with the sequences of the mammalian enzymes (37). Major regions of homology within the two subtilisins included the "active serine" sequences and the areas containing the histidine sequences. Since the catalytic mechanism of the subtilisins

is assumed to be the same as that of the rest of the serine proteases, it is likely that only one histidine is involved. As yet there is no evidence to suggest which one it might be. Recently other bacterial proteases and their "active serine" sequences have been reported (45-47). The total evidence suggests that the bacterial enzymes evolved independently of the mammalian proteases (37). This will be discussed at greater length later in this thesis.

Sorangium sp. α-lytic protease

The isolation in pure form of the serine proteases α and β-lytic protease from the bacterium Sorangium sp. by Whitaker (1) among other things helped to answer the question of whether one or two histidines were involved in the mechanism of catalysis by the serine proteases. The α enzyme, which apparently operated by the same mechanism as the mammalian serine proteases contained only one histidine, thus strongly supporting the necessity of only one histidine in the mechanism. Kinetic comparisons of chymotrypsin and α-lytic protease presented by Whitaker and Kaplan (36) suggested that both enzymes operated by the same mechanism. The pH dependence of the catalytic rate constants was the same for both enzymes and the pK value of 6.7 accompanied by a shift to 7.35 when water was replaced by D_2O were also common properties of both proteases. This was consistent with the requirement for a single unprotonated imidazole group and showed that the catalytic mechanism need involve only one histidine.

Other properties of α-lytic protease have been determined

by Whitaker and co-workers (1,29,36,51-55). The sedimentation coefficient was determined as 2.2 Svedberg units and the ultracentrifuge pattern showed one peak. The partial specific volume using Cohn and Edsall's method was estimated to be 0.72. The α enzyme was found to consist of 198 amino acid residues and the molecular weight was determined as 20,100 using a statistical method for computing a "best estimate" of the multiplier which converts composition per unit weight of enzyme preparation to composition per mole of enzyme (55). The molecular weight that had previously been estimated by the Archibald method was 19,000 (52). A series of experiments showed that generally the linkages split by the enzyme involve the carboxyl group of a neutral, aliphatic amino acid. The α enzyme appears to be metal free and is readily inactivated by DFP or sarin.

It is interesting to note that although α-lytic protease is a bacterial enzyme, it has much in common with the mammalian serine proteases. From Tables 1-3 and 1-4 it can be seen that in α-lytic protease the "active serine" sequence, a disulfide bridge, and the histidine sequence are very similar to the mammalian counterparts. For this reason there has been much interest in comparing the sequences of critical portions of this molecule to those of the other proteases (2). If this enzyme were structurally similar to the mammalian serine proteases, it would be the first bacterial proteolytic enzyme to display such a resemblance. It is readily seen that the first step in attempting to draw structural similarities and thus evolutionary suggestions is to obtain the complete amino acid sequence of the protein in question. To this end, this thesis

attempts to contribute. In a later chapter more will be said about the possible evolutionary significance of α-lytic protease.

CHAPTER II

PEPTIDES RESULTING FROM A TRYPTIC DIGEST OF S-CARBOXYMETHYLATED α-LYTIC PROTEASE

1. Introduction

Elucidation of the complete amino acid sequence of a protein requires the results of several enzymatic digests since no one approach can give the necessary overlapping sequences. A common order of methods is to first sequence the peptides resulting from digestion by an enzyme. Hydrolyzing with a second enzyme produces cleavages at different points on the molecule, thus yielding other peptides, some of which will be overlapping sequences of fragments obtained from the first digest. The segments can then be fitted together to form longer sequences.

Before the present work was begun, certain parts of the α-lytic protease molecule had already been sequenced (26). A peptic digest of the protein had been subjected to the diagonal electrophoretic technique of Brown and Hartley (56) resulting in isolation and sequence determination of five of the six cysteic acid peptides. Only an amino acid composition of the sixth peptide was obtained since it was isolated in too low a yield to allow sequence elucidation. Some of these cysteic acid peptides were later isolated from the digests described in this thesis. All previously isolated peptides are presented in Table 2-1.

The choice of trypsin as the degrading agent for the continuation of the sequence elucidation of α-lytic protease was made because of its several convenient properties. Trypsin

Table 2-1

Previously Isolated Peptides of α-Lytic Protease

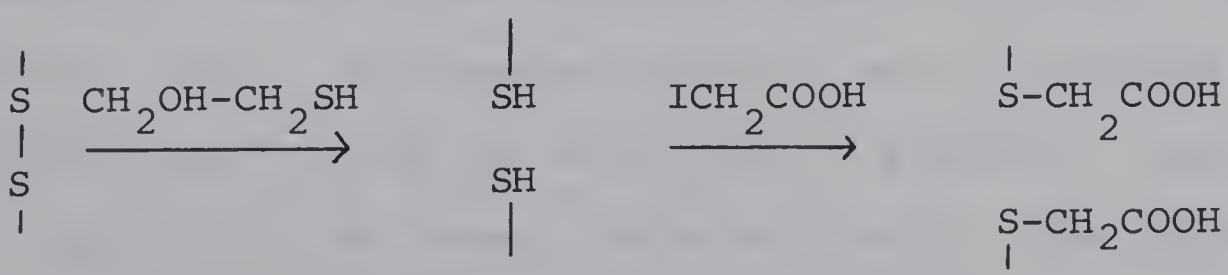
Peptide	Composition or Sequence
CDPB1a CDPB1b	Phe Val Thr Ala Gly His Cys Gly Thr Val Asn Ala S S
CDPB2	Cys Ser Val Gly Phe
CDPD2	Ala Ala Val Cys Arg Ser Gly Arg Thr Thr Gly Tyr Gln Cys Gly Thr Ile Thr S S
CDPFTB2	Asn Ala Cys Met Gly Arg S S
CDPFB5	(Asx Asx Cys Gly)
CDPFB4	(Asx Asx Cys Gly Gly Pro Ile)
CDPFTB3a	(Asx Asx Cys Gly Gly Pro Ile Asx Ser Ser Glx Glx Ala Val Asx Ser Gly Gly)
CDPFTB3b	(Asx Asx Cys Gly Gly Pro Ile Asx Ser Ser Glx Glx Ala Val)
CDPFTB3c	(Asx Asx Cys Gly Gly Pro Ile Asx Ser Ser Glx Glx Ala Val Asx)

is very specific in its action and therefore few peptides should be obtained which result from partial splitting of bonds or cleavages at sites other than basic residues. The peptides should also be isolable in good yields. However, since trypsin only hydrolyzes sites involving the carboxyl group of basic residues, especially if it is inhibited to minimize any inherent chymotryptic activity, portions of the molecule will remain largely undegraded if the basic residues are not fairly uniformly distributed throughout the molecule. Extreme difficulties are often encountered with these large fragments since they are usually insoluble in water and do not move well when subjected to high voltage electrophoresis, the principal tool used for peptide purification in this study. Thus although results of hydrolysis by trypsin cannot in themselves determine the order of the peptides within the molecule, tryptic digestion is a logical starting procedure for determining the amino acid sequence of a protein.

To help unfold the protein and therefore assist approach of the degrading enzyme, the disulfide bridges were reduced and the resulting sulfhydryl groups derivatized. The S-carboxymethylated protein as a choice of derivative is briefly discussed at the end of this chapter.

2. Preparation of S-carboxymethylated α-lytic protease

Basically the reaction sequence is as follows:



100 mg of α-lytic protease was dissolved in 10 ml of 0.1 M tris-acetate buffer, pH 8.0, at 5°C and 100 µl of 1 M DFP (Mann Analyzed) was added. The solution was left at 5°C for 2 hours to convert the enzyme to the inactive DFP derivative. The pH was then adjusted to 3.0 with 6 M HCl using a Radiometer type TTTla pH meter. 6 g of recrystallized urea was added and the solution was allowed to reach room temperature. The pH was readjusted to exactly 3.0 and the solution left at room temperature for 30 minutes to assure complete denaturation.

200 µl of 2-mercaptoethanol (Eastman) was then added and the pH raised to 8.0 with 6 M NH_4OH. The tube was flushed with nitrogen, capped and incubated at 37°C for 4 hours. After incubation the solution was transferred under nitrogen to a centrifuge tube and 100 ml of a mixture of 98% ethanol - 2% conc. hydrochloric acid (v/v) was added. A fine precipitate of reduced protein was produced and left to develop at -20°C overnight, then centrifuged at 13,000 x g for 1 hour in an International refrigerated centrifuge. The precipitated protein was suspended in 10 ml of an 8 M urea solution, pH 2, and 100 mg of iodoacetic acid (recrystallized from petroleum ether) was added. The pH was then raised to 8.6 with 6 M NH_4OH (at which point the protein partially dissolved) and maintained at this level by the addition of dilute NH_4OH (constantly keeping the solution under nitrogen). After 30 minutes, 600 µl of 2-mercaptoethanol was introduced and the pH maintained at 8.6 for a further 15 minutes. This assured the destruction of excess iodoacetic acid. The pH was then

lowered to 3.0 by the addition of 6 M HCl and the solution dialyzed against 10^{-3}M HCl. The suspension of precipitated S-carboxymethylated protein was freeze-dried. The yield was 75 mg and amino acid analysis on a Beckman model 120 B amino acid analyzer showed 5.9 residues of S-β-carboxymethylcysteine were obtained (theoretical = 6). The recovery of histidine was 0.86 residues (theoretical = 1) and the yield of methionine was calculated to be 1.85 residues (theoretical = 2).

3. Digestion with trypsin

100 mg of carboxymethylated α-lytic protease was weighed into a pH stat tube and dissolved as much as possible in 20 ml of 0.05 M NH_4OH. The pH was adjusted to exactly 8.0 using a Radiometer type TTT1a automatic titrator. The volume of titrant was measured with a Radiometer SBR 2c Titrigraph recorder. The temperature was 25°C and the titrant 0.10 M NaOH. When pH 8.0 was attained, no further base uptake was observed for 10 minutes. Then 250 µl of a TPCK inhibited trypsin solution (10 mg of TPCK trypsin dissolved in 1.25 ml of 10^{-3}M HCl) was added and the solution left in the pH stat apparatus for 5 hours. The suspension was then centrifuged on an International clinical centrifuge for 15 minutes to separate the remaining precipitate and the supernatent was applied on electrophoresis paper immediately to prevent further hydrolysis (see part 3 of this chapter). From this time on, the digest was treated as 2 parts, soluble and insoluble fractions. The precipitate was washed, freeze-dried and weighed and accounted for approximately 1/6 of the protein material digested.

Assuming the pKa of the amino groups to be 7.5, a corrected

calculation for the number of groups titrated per protein molecule, based on the number of moles of NaOH consumed during digestion in maintaining the pH at 8.0, showed that 11.7 peptide bonds had been cleaved per mole of protein. Theoretically 14 bonds should have been hydrolyzed. Considering the assumptions involved in this calculation, particularly in the pKa value, the agreement is not unsatisfactory and indicates that the tryptic hydrolysis was essentially complete.

4. Isolation, purification and sequence elucidation

Only the soluble portion of the digest was utilized in the present study. As previously mentioned, the digest supernatant was applied at a level of 0.07 μmoles per cm on Whatman 3MM filter paper immediately after centrifugation to prevent further hydrolysis. The paper was wetted with pH 6.5 buffer (composition 879 ml H_2O, 100 ml pyridine and 3 ml glacial acetic acid) and subjected to electrophoresis at 3 Kv for 40 minutes. For complete details of the apparatus and procedure the reader is referred to the thesis of K. Stevenson (24). The resulting peptide bands were detected by the staining of side strips with the cadmium-ninhydrin reagent described elsewhere (24).

The peptides were designated "T" peptides (tryptic) and numbered with respect to decreasing basicity, T_1 having the highest mobility towards the negative electrode. All Tn peptides resulted from a peptide band which was neutral at pH 6.5 and was separated and purified by further treatment. The following is a list of bands resulting from the original pH 6.5 electropherogram: T1, T2, T3-7 (so named because it

appeared to be four peptides very close together), T8-9, T10, Tn and T11-12. For a complete list of all tryptic peptides eventually purified, see Table 2-2. Peptides were subjected to amino acid analysis after acid hydrolysis in constant boiling HCl for 16 - 20 hours at 110°C on a Beckman model 120 B amino acid analyzer. All electrophoresis was done at 3 Kilovolts (3000 volts). N-terminal analyses were routinely done when a pure peptide was isolated by utilizing the "Dansylation" procedure as outlined by Stevenson (24).

Basic and acidic peptides

T1 and T2

These peptides were isolated in very small amounts and were ignored in the present study.

The T3-7 region

The T3-7 region was subjected to further electrophoresis at pH 1.8 (buffer composition of 2% formic acid, 8% acetic acid and 90% water) for 45 minutes, which produced 2 bands after cadmium-ninhydrin development, T3-7a and T3-7b, which upon amino acid analysis proved to be still impure. T3-7a was finally purified by electrophoresis at pH 6.5 for 1 hour producing T3-7a1 and T3-7a2. T3-7b was purified by electrophoresis at pH 3.5 (buffer composition 1890 ml H_2O, 10 ml pyridine, and 100 ml glacial acetic acid) for 1 hour. T3-7b1 and several bands of free amino acid contamination resulted. T3-7a1, T3-7a2 and T3-7b1 were then sequenced by the "Dansyl-Edman" procedure as outlined by Gray (66).

The final sequences and molar ratios of amino acids were as follows:

T3-7a1	Ala →	Ala →	Arg →	
	1.06	1.06	0.88	
T3-7a2	Ser →	Gly →	Arg →	
	0.96	1.10	0.91	
T3-7b1	Gly →	Ala →	Thr →	Lys →
	0.94	1.06	1.04	1.00

In these and other peptides mentioned in this thesis, an arrow under the amino acid indicates that the residue has been successfully determined by the "Dansyl-Edman" procedure. No arrow indicates that the sequence of that particular amino acid was determined by other workers. Parentheses around a group of amino acids are used to indicate that the sequence is unknown.

T8-9

The T8-9 band was purified by electrophoresis at pH 1.8 for 1 hour. T8 was not recovered in adequate yield for characterization. Since T9 was isolated in low yield it was decided to proceed with the "Dansyl-Edman" treatment even though the peptide was not completely pure. In the sequence elucidation serine was present in small amounts at each step and may be only an impurity. Due to a lack of sufficient peptide, the sequence was not completed and several of the steps were questionable. However, recently in this laboratory a portion of T9 has been isolated from a chymotryptic

digest and partially sequenced (57). This peptide confirmed the partial elucidation in the present work. A peptide with a composition identical to that of T9 was isolated from a tryptic digest of S-aminoethylated protein, but this fragment was not obtained in a sufficient quantity for further study (57). The following composition and sequence are suggested.

T9	Ile →	Gly →	Gly →	Ala →	Val →	Val →	(Gly	Thr	Phe	Ala	Ala	Arg)	also Ser
	0.77	0.92	0.92	0.90	1.03	1.03	0.92	1.06	1.00	0.90	0.90	0.88	0.59

T10

T10 was purified by electrophoresis at pH 1.8 for 1 hour. The amino acid analyses (20 hour hydrolysis and 70 hour hydrolysis results) supported the suspicion that this was an extended sequence of the histidine peptide previously isolated and sequenced (2). The previously elucidated peptide had an N-terminal phenylalanine residue and T10 had an N-terminal sequence determined by the "Dansyl-Edman" method as Gly Phe-. From previous work it was known that a peptic digest of T10 should release a tripeptide (Thr Ala Arg) if it was an extension of the known fragment. The peptide T10 was therefore digested with pepsin (Worthington, 2X recrystallized) using a 50:1 protein:enzyme ratio for 5 hours at 37°C, and the fragments purified by electrophoresis at pH 6.5 and pH 1.8. This resulted in a series of peptides. Upon analysis T10P2 (the second most basic peptic fragment of T10) was found to have the sequence Thr Ala Arg confirming that the peptide T10 was in fact an extended sequence of the previously determined histidine peptide CDPBla (see Table 2-1).

T10

Gly Phe Val Thr Ala Gly His Cys Gly Thr Val Asn Ala Thr Ala Arg

0.98 1.00 0.91 0.94 1.01 0.98 1.11 0.65 0.98 0.94 0.91 1.05 1.01 0.94 1.01 1.00

T10P2

0.98 1.12 0.94

The T11-12 region

The T11-12 region was separated and purified by electrophoresis at pH 1.8 for 80 minutes. The two resulting peptides were T11 and T12. The amino acid composition of T11 suggested that a peptic digest would produce smaller, more easily workable fragments, so the peptide was incubated with pepsin under the same conditions as outlined earlier. Of the many fragments that resulted, only three peptides were obtained in good yield and these accounted for the total amino acid composition of T11. The "Dansyl-Edman" method provided the sequence of the three peptides: T11P2 (Gly Ser Thr Glu), T11Pn4 (Ala Ala Val Gly) and T11Pn1 (Ala Ala Val Cys Arg). The last was a previously sequenced peptide (CDPD2 in Table 2-1). From the knowledge that the N-terminal residue of T11 was glycine, it was obvious that T11P2 must be in the N-terminal portion. Since T11 resulted from a tryptic digest, the C-terminal residue would most likely be a basic one. C-terminal arginine in T11Pn1 fitted this requirement. This left the fragment T11Pn4 necessarily in the middle of the peptide T11. Following is the total sequence and composition of the original peptide and the digest peptides. The elucidation of this sequence clearly provided an extension of the disulfide structure

previously determined.

T11	Gly	Ser	Thr	Glu	Ala	Ala	Val	Gly	Ala	Ala	Val	Cys	Arg
	0.98	0.88	0.98	1.04	1.01	1.01	0.95	0.98	1.01	1.01	0.95	0.88	0.90
			P2			Pn4				Pn1			
	1.00	0.86	1.02	1.01	1.02	1.02	1.00	1.00	1.00	1.00	1.00	0.71	0.94

T12 was not isolated in a large enough yield to allow a sequence determination. A peptic digest was done and a small peptic fragment, T12P1, was isolated. The rest of the sequence shown was elucidated by another worker in this laboratory (58) who isolated the same peptide from a tryptic digest of S-aminoethylated protein. Below is the sequence and composition of this peptide.

T12	Ala	Asn	Ile	Val	Gly	Gly	Glu	Ile	Tyr
	1.09	1.04	0.75	0.87	1.00	1.00	1.04	0.75	0.50
							P1		
							1.04	0.92	0.49

The neutral peptides

The neutral region of the original pH 6.5 electropherogram, Tn, was separated by electrophoresis at pH 6.5 for 6 hours. This resulted in seven bands visible after staining with the cadmium-ninhydrin reagent. In order of decreasing basicity they were the Tn1-3 region, Tn4, Tn5, Tn6, Tn7, Tn8 and Tn9.

The Tn1-3 region

This region was further purified by electrophoresis at pH 1.8 for 1 hour producing a series of peptides of which only two were isolated in workable amounts. Tn1-3d was sequenced

by another worker in this laboratory (58) and Tn1-3f, which was impure after electrophoresis at pH 1.8, was subjected to electrophoresis at pH 3.5 for 2 hours. Tn1-3f1 and several weakly staining bands resulted. Only Tn1-3f1 was recovered in adequate amounts for characterization. Since this peptide was an extension of the previously isolated peptide CDPFTB2 (Table 2-1) the sequence was determined by the "Dansyl-Edman" method only far enough to give a conclusive result.

Tn1-3d	Val	Phe	Pro	Gly	Asn	Asp	Arg
	0.88	0.95	0.99	1.00	1.02	1.02	1.05

Tn1-3f1	Gly→	Leu→	Thr→	Gln→	Gly→	Asn→	Ala→	Cys	Met	Gly	Arg
	1.00	0.96	1.07	1.09	1.00	1.13	1.04	0.88	0.82	1.00	0.96

Tn4

Tn4 was purified by electrophoresis at pH 1.8 for 1 hour and was sequenced by the "Dansyl-Edman" method. However, an uncertainty demanded that further study be done. A troublesome N-terminal tyrosine residue forced the employment of a peptic digest under the same conditions used previously. Purification of the fragments by electrophoresis at pH 6.5 and pH 1.8 produced a peptide which conclusively gave an N-terminal tyrosine upon dansyl chloride treatment. Following are the sequences of the peptide Tn4 and a peptic fragment Tn4P2.

Tn4	Tyr→	Ala→	Glu→	Gly→	Ala→	Val→	Arg→
	0.33	1.00	1.04	1.05	1.00	1.05	1.05
	⟵→——P2——						
	0.73	0.93	1.04	1.00			

Tn5

Tn5 was purified by electrophoresis at pH 1.8 for 1 hour and the sequence was determined by others in this laboratory (58). The composition and sequence are presented below.

Tn5	Ser→	Ser	Leu	Phe	Glu	Arg
	0.94	0.94	1.00	1.06	1.06	1.00

Tn6

Tn6, purified by pH 1.8 electrophoresis for 1 hour, was strongly suspected of being an extended sequence of a previously determined peptide CDPD2. Since CDPD2 (see Table 2-1) had been the result of a peptic digest, Tn6 was subjected to pepsin hydrolysis, under the same conditions as outlined earlier, in an attempt to isolate the extending portion. Of the fragments produced, Tn6P1 proved to be the extending portion -Ala Lys and Tn6P4 was a section of CDPD2 with the extending portion at the C-terminal end, thus verifying the total sequence shown below. Other segments of Tn6 were also found but are not vital to the extension of the previously elucidated peptide.

Tn6	Thr→	Thr	Gly	Tyr	Gln	Cys	Gly	Thr	Ile	Thr	Ala	Lys
	0.94	0.94	0.98	0.40	1.03	0.70	0.98	0.94	0.91	0.94	1.00	0.91
										⟵P4⟶		
										0.88	1.00	1.06
											⟵P1⟶	
											1.00	1.02

Tn7

Tn7 was further purified by electrophoresis at pH 1.8 for

1 hour producing Tn7a and Tn7b. Only Tn7a was recovered in adequate amounts for further characterization. An attempt was made to sequence the peptide using the "Dansyl-Edman" method but a lack of sufficient material prevented its completion. The sequence has since been completed by others (58) and is presented below.

Tn7a	Asn→	Val→	Thr→	Ala→	Asn→	Tyr→	Ala→	Glu→	Gly	Ala	Val	Arg
	0.91	0.93	0.93	0.99	0.91	0.78	0.99	1.00	1.10	0.99	0.93	0.88

Tn8

The peptide Tn8 was not isolated in a sufficient amount to allow even a satisfactory amino acid analysis.

Tn9

Tn9 was separated from other bands well enough on the 6 hour, pH 6.5 electropherogram to be eluted directly from it. The sequence was determined by the "Dansyl-Edman" procedure but the last residue could not be verified due to a failure of the Edman step at the asparagine residue. Asparagine has previously been reported to sometimes cyclize into an imide which opens to give predominantly a β-aspartyl peptide (59). This β bond does not undergo cleavage in the cyclization step of the Edman degradation. Since the peptide is neutral at pH 6.5 the aspartic acid residues must exist in the amide form. It is also interesting to note that this peptide resulted from an hydrolysis at an alanine residue. Since the peptide was isolated from a tryptic digest it is assumed that it arose from an autolytic cleavage either during preparation

Table 2-2

Peptides from the Trypsin Digest of

Amino Acid

Peptide	Lys	His	Arg	Asp	Thr	Ser	Glu	Pro	Gly	Ala	CMCys[4]	Val
T3-7a1			0.88							2.12		
T3-7a2			0.91			0.96			1.10			
T3-7b1	1.00				1.04				0.94	1.06		
T9			0.88		1.06	0.59[3]			2.76	2.70		2.06
T10		1.11	1.00	1.05	2.82				2.95	3.08	0.65	1.92
T11			0.90		0.98	0.88	1.04		1.96	4.04	0.88	1.90
T12				1.04			1.04		2.00	1.09		0.87
Tn1-3d			1.05	2.04				0.99	1.00			0.88
Tn1-3f1			0.96	1.13	1.07		1.09		3.00	1.04	0.88	
Tn4			1.05				1.00		1.05	2.00		1.05
Tn5			1.00			1.88	1.06					
Tn6	0.91				3.76		1.03		1.96	1.00	0.70	
Tn7a			0.88	1.82	0.93		1.00		1.10	2.98		1.85
Tn9				1.80	1.05					1.10		1.00
Total of each amino acid	2	1	10	9	13	4	7	1	20	22	4	12

[1] calculated with respect to lysine
[2] calculated with respect to aspartic acid
[3] probably impurity
[4] S-β-carboxymethylcsyteine

S-Carboxymethylated α-Lytic Protease

Composition					Mobility at pH 6.5	Cadmium Ninhydrin Color	% Yield	Other Comments
Met	Ile	Leu	Tyr	Phe				
					0.42[1]	red	4.0	
					0.42[1]	orange	15.7	
					0.42[1]	orange	10.0	
	0.77			1.00	0.23[1]	red	5.7	
				1.00	0.16[1]	yellow	4.0	Stains for his
					0.23[2]	yellow	10.0	
	1.50		0.50		0.28[2]	red	4.3	Stains for tyr
				0.95	0.00	red	9.2	
0.82		0.96			0.00	yellow	4.3	
			0.33		0.00	red	4.0	Stains for tyr
		1.00		1.06	0.00	orange	9.5	
	0.91		0.40		0.00	red	9.5	Stains for tyr
			0.78		0.00	yellow-orange	3.7	Stains for tyr
					0.00	yellow	4.3	
1	4	2	4	4	Total amino acids accounted for = 120 residues			

of the enzyme or in solution before inactivation by DFP. The possibility that it represented the C-terminal end of the protein has been ruled out by the demonstration by others that it is derived from an interior portion of the polypeptide chain (see Table 4-1).

Tn9	Asn→	Val→	Thr→	Asn→	Ala
	0.82	0.91	0.95	0.82	1.00

5. Discussion

It is interesting to note that although TPCK treated trypsin was used to minimize if not eliminate chymotryptic splits, two peptides appeared which were not the result of hydrolysis at a basic residue. Tl2, which originated by cleavage at a tyrosine residue, seemed to be the result of a chymotryptic hydrolysis. Tn9, as previously mentioned, was more unexpected under the circumstances and was probably the result of autolysis during purification or during the preparation of the S-carboxymethylated derivative. However, neither peptide was isolated in good yield (see Table 2-2).

Although the tryptic digest did produce a number of peptides suitable for sequence analysis, a good portion of the protein remained in the insoluble portion of the digest. This could have been due to either the insolubility of the reduced and S-carboxymethylated protein, which prevented large areas of the molecule from coming into contact with the hydrolysing enzyme, or to large areas of the protein that are void of basic residues and thus are immune to the action of trypsin. Whatever the cause, it is apparent from

Table 2-2 that only 120 of the 198 amino acid residues of α-lytic protease could be accounted for, and any suggestions regarding the sequence of large portions of the molecule would have to await further study.

CHAPTER III

PEPTIDES FROM A CHYMOTRYPTIC DIGEST OF S-AMINOETHYLATED α-LYTIC PROTEASE

1. Introduction

The production of a large proportion of insoluble "core" material during the tryptic digestion of the S-carboxymethylated α-lytic protease prompted a reassessment of the approaches being employed in the elucidation of the amino acid sequence of this protein. It was reasoned that the conversion of all the cystine residues into S-β-aminoethyl derivatives would provide six extra charges on the protein and perhaps increase the solubility of the enzyme. Since the structure of S-β-aminoethylcysteine resembles that of lysine, this derivative would also provide six additional cleavage points for trypsin. The sequence around five of the six half-cystines was known, so hydrolysis at these points would not cleave fragments that were potential overlapping sequences, the only exception being the half-cystine whose sequence had not been determined.

Since a tryptic digestion of the S-aminoethylated enzyme was being done in this laboratory (58) and looked very promising, a chymotryptic digest was performed in an attempt to isolate peptides which could not be liberated by the action of trypsin. It was also hoped that chymotrypsin would cleave the polypeptide chain into different fragments of the same area that yielded peptides which had been sequenced earlier. This would provide overlapping structures of the tryptic peptides isolated previously from both the digestion of the S-aminoethylated enzyme by Dr. N. Nagabhushan and the hydrolysis

of the S-carboxymethylated protein described in Chapter II.

2. Reduction and aminoethylation

The procedures used to reduce and aminoethylate the protein were basically those of Raftery and Cole (60, 61). 200 mg (10 μmoles) of α-lytic protease was dissolved in 20 ml of 0.1 M tris buffer, pH 8.0, at 5°C and 200 μl of a 1 M DFP solution was added. The solution was left at 5°C for two hours to ensure complete inactivation, then dialysed against distilled water overnight and freeze-dried.

The lyophilized material was then dissolved in 20 ml of a 1 M ammediol buffered 8 M urea solution, pH 3.0, prepared by dissolving 9.6 g (160 mmoles) of ultra pure urea (Mann Research Laboratories Inc.), 2.1 g (20 mmoles) of ammediol (2-amino-2-methyl-1,3-propanediol, Eastman) and 2 mg (5.4 μmoles) of EDTA (disodium salt) in approximately 15 ml of deionized water, bringing the pH to 3.0 with conc. HCl and then diluting to a volume of 20 ml with deionized water. After leaving the solution at room temperature for a half hour to ensure complete denaturation, 400 μl (5.76 mmoles) of 2-mercaptoethanol was added, the pH was raised to 8.0 with 6 M NH_4OH and the tube was flushed with nitrogen and capped. This solution was left at 37°C for 4 hours.

After the reduction was complete, a total of 3 ml (58.0 mmoles) of ethyleneimine (Dow Chemical) was added, taking care to keep the solution under a nitrogen atmosphere and the pH below 9.0. The ethyleneimine was added as separate 1 ml aliquots at 15 minute intervals. The total reaction time in the presence of ethyleneimine was 1 hour (the solution stood

for a half hour after the last addition of ethyleneimine). The pH was then lowered to 5.0 since this procedure had been reported to produce better yields of S-β-aminoethylcysteine (62), and the solution was dialyzed thoroughly against distilled water, then lyophilized. The yield of S-aminoethylated α-lytic protease was 180.6 mg with the amino acid analysis showing 3.8 residues of S-β-aminoethylcysteine (theoretical = 6), 1.3 residues of histidine (theoretical = 1) and 1.0 residues of methionine (theoretical = 2). The low conversion to the S-aminoethyl derivative is discussed at the end of this chapter.

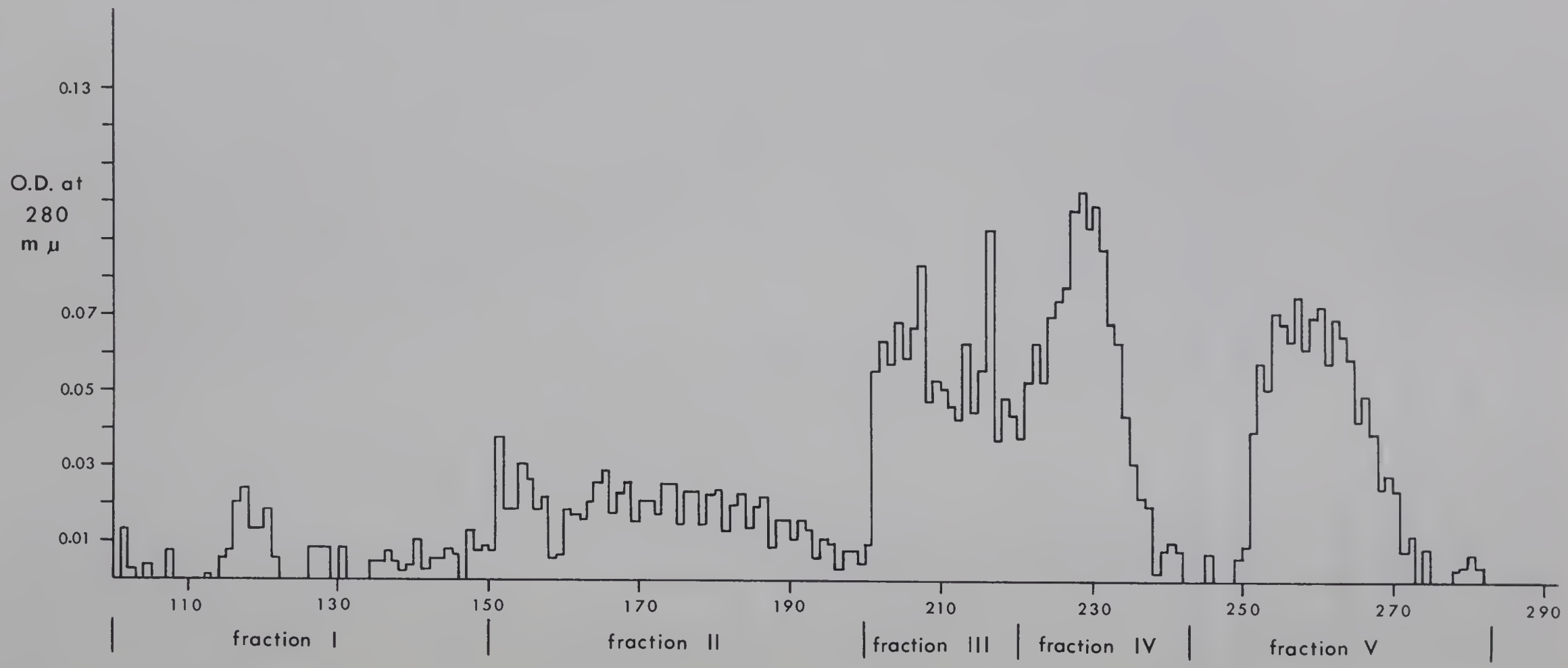
The hydrolysis procedure for the amino acid analysis of the aminoethylated protein was slightly different from the usual technique described elsewhere (24). A special evacuation technique and twice distilled, constant boiling HCl were employed. The protein was dissolved in the constant boiling HCl inside of a large test tube and the tube was pulled to a capillary. The solution inside the tube was frozen using a dry ice - acetone bath and the tube was evacuated to 30 microns pressure. The solution was then allowed to melt (under evacuation) and the tube evacuated until no further foaming or bubbles appeared. The solution was refrozen and again allowed to melt (still at 30 microns pressure). After melting the second time evacuation was continued for an additional 10 minutes at 30 microns pressure or lower, and the tube was sealed and hydrolyzed for 20 hours at 110°C. Using this method the yields of S-β-aminoethylcysteine recovered from hydrolysis were consistently 95% (based on hydrolysis of a

known amount of S-β-aminoethyl-L-cysteine). It was thus concluded that the low yields of S-β-aminoethylcysteine recovered from the protein hydrolysates were not due to destruction during hydrolysis.

3. Digestion of S-aminoethylated α-lytic protease with chymotrypsin and fractionation on Sephadex G-25

100 mg of S-aminoethylated α-lytic protease were dissolved in 15 ml of deionized water (brought to pH 8.0 with dilute NH_4OH) and 0.1 μmoles of α-chymotrypsin dissolved in dilute NH_4OH, pH 8.0, was added, giving a protein:enzyme ratio of 50:1. This solution was left at 37°C for 5 hours with periodic adjustment of pH to 8.0 using dilute NH_4OH, then centrifuged on an International clinical centrifuge. No sediment was observed. The solution was then applied to a Sephadex G-25 column (4.3 cm x 195 cm) and eluted with 0.05 M acetic acid. 10 ml fractions were collected and the optical density at 280 mμ was measured on a Beckman DU spectrophotometer. On the basis of the optical density the solutions were categorized into five fractions. The elution profile and division into fractions is shown in Figure 3-1. The yield of material with an optical density eluted from the column was, within experimental error, essentially 100%. Each fraction was lyophilized and redissolved in a smaller amount of deionized water, then applied on Whatman #1 paper and subjected to electrophoresis as outlined below. Only fractions II, III, IV and V were found to contain peptide material.

Figure 3-1 Elution profile of the Sephadex G-25 separation



4. Isolation, purification and sequence elucidation

In many cases the peptides were isolated in low yield. Since this ruled out the possibility of cleaving these fragments into smaller pieces by treatment with other proteolytic enzymes, the "Dansyl-Edman" procedure was used on such peptides. As many residues as possible were determined until the material was exhausted or a conclusive sequence had been determined. This was found to produce the most satisfactory results in such cases. Any doubtful sequence result will be mentioned. Most of the peptides isolated from this digest were suspected of being segments or extensions of peptides already sequenced or overlapping peptides between two known sequences. Because of this it was unnecessary in many cases to completely elucidate the sequence. For a more complete picture of the overlapping peptides than can be obtained from the results in this chapter, the reader is referred to Table 4-1 in Chapter IV.

(a) Fraction II

Preliminary results of electropherograms of Fraction II were not encouraging since considerable adsorption and trailing of the peptides appeared to be occuring. It was considered that the fraction was composed of large peptides which would be difficult to isolate and purify by paper methods. For this reason the fractionation of this material was not attempted in the present work. However, Dr. M. Olson in this laboratory has subsequently been successful in purifying the major peptides of this fraction and his results are included in Table 4-1 of Chapter IV.

(b) Fraction III

Fraction III was subjected to electrophoresis at pH 6.5 for 80 minutes which produced the following bands after cadmium-ninhydrin staining: CIII1-2, CIII3, CIII4, CIII5-6, CIII7-8, CIII9, CIII10, CIII11, CIII12, CIIIn, CIII13-14 and CIII15-16.

The CIII1-2 region

CIII1-2 was separated by electrophoresis at pH 1.8 for 45 minutes producing CIII1 and CIII2. CIII2 was not recovered in adequate amounts for further purification. CIII1 was suspected of being an overlap between T11 and T3-7a2 sequenced previously. The "Dansyl-Edman" procedure was employed to confirm this. The sequence and molar ratio of amino acids are presented below.

CIII1	Arg→	Ser→	Gly	Arg
	0.92	0.91	1.25	0.92

CIII3

CIII3 was purified by electrophoresis at pH 1.8 for 50 minutes resulting in two peptides called CIII3a and CIII3b. CIII3a was further purified by electrophoresis at pH 3.5 for 50 minutes. CIII3b was suspected of being part of the peptide T3-7b1 and the N-terminal portion of T10 which were both discussed earlier. Sequence analysis by the "Dansyl-Edman" procedure confirmed this supposition. CIII3a was almost certainly the same as T3-7a1 discussed in Chapter II. The results for both peptides are shown below.

CIII3a	Ala→	Ala	Arg
	0.89	0.89	1.04

CIII3b	Arg→	Gly→	Ala→	Thr→	Lys	Gly	Phe
	0.70	1.16	1.08	1.13	0.97	1.16	0.84

CIII4

CIII4 was not recovered in adequate amounts for further characterization.

The CIII5-6 region

This region was separated by electrophoresis at pH 1.8 for 50 minutes producing three peptides, CIII5-6a, CIII5-6b and CIII5-6c. Further attempts at the purification of CIII5-6b and CIII5-6c were unsuccessful and no additional information is available concerning these bands. The composition and N-terminal analysis of CIII5-6a indicated that it was derived from the C-terminus of Tn1-3f1 previously described.

CIII5-6a	Gly→	Arg
	1.1	0.91

The CIII7-8 region

The CIII7-8 region was separated by electrophoresis at pH 1.8 for 1 hour producing CIII7-8a and CIII7-8b. CIII7-8b required further purification by electrophoresis at pH 3.5 for 50 minutes. Both peptides were suspected of being portions of known sequences, CIII7-8a being the C-terminus of τIV5-6c1 and CIII7-8b the C-terminus of τIV5-6a (see Table 4-1). Although the purity of these peptides was not totally adequate, it seemed to be sufficient for characterization. Since CIII7-8a is a basic peptide, the glutamic acid residue must exist as the amide.

CIII7-8a	Ser→	Gln→	Arg→				
	0.77	1.15	1.04				
CIII7-8b	Val→	Thr	Arg				
	1.10	0.82	1.13				

CIII9

CIII9 was purified by pH 1.8 electrophoresis for 40 minutes and is part of the previously isolated histidine sequence CDPB1a (see Table 2-1).

CIII9	Val→	Thr	Ala	Gly	His
	0.88	0.93	1.04	1.00	0.91

CIII10

CIII10 was purified by electrophoresis at pH 1.8 for 40 minutes and provided a good overlap between two previously isolated peptides Tn7a and Tn6 (see Table 4-1). The "Dansyl-Edman" procedure was continued for four steps to confirm this assignment.

CIII10	Ala→	Lys→	Asn→	Val→	Thr	Ala	Asn	Tyr
	0.90	1.07	1.01	1.01	0.96	0.90	1.01	0.82

CIII11

This peptide was subjected to pH 1.8 electrophoresis for 40 minutes for purification and is clearly a part of Tn1-3f1 sequenced earlier.

CIII11	Thr→	Gln	Gly	Asn	Ala	Cys	Met
	0.82	0.89	1.03	1.10	1.07	0.39	0.63

CIII12

This peptide, purified by pH 1.8 electrophoresis and repurified by electrophoresis at pH 3.5 was not obtained in an adequate state of purity. Since this fraction contained S-β-aminoethylcysteine an attempt was made to purify the fragments resulting from a tryptic digestion. However, this also failed to yield peptides which could be purified. It is suspected that this band is actually two closely related peptides that have the same mobility under the conditions of fractionation.

The CIII13-14 region

This region proved to be quite complex. Upon electrophoresis at pH 1.8 for 80 minutes, four ninhydrin staining bands resulted. CIII13-14d was satisfactorily pure after this treatment but CIII13-14b and CIII13-14c were further purified by electrophoresis at pH 3.5 for 2.5 hours and 50 minutes respectively. This resulted in two peptides, CIII13-14b1 and CIII13-14b2. CIII13-14c was not isolated in a pure form. CIII13-14a was also not obtained in a satisfactory state of purity. It is suspected that these bands are each two peptides with the same electrophoretic mobility. CIII13-14b2 was part of the known peptide CDPD2 (Table 2-1) and CIII13-14b1 was lost during purification. Only an N-terminal analysis and amino acid analysis were obtained for this peptide.

CIII13-14d was suspected of being part of a known peptide τIV7-8ab1. To confirm this an α-lytic protease digest was performed on the peptide by dissolving it in N-ethyl morpholine buffer, pH 8.0, and adding α-lytic protease dissolved in N-ethyl morpholine buffer (50:1 peptide:enzyme ratio). The

solution was incubated at 37°C for 5 hours, then purified by electrophoresis at pH 6.5 and 1.8. CIII13-14d and the fragments which confirmed its sequence are shown below. Since the peptide CIII13-14dLn2 contains both aspartic acid residues and is neutral, these residues must exist in the amide form.

CIII13-14b1

Ser→ (Leu Gly Thr Val)

1.12 0.97 0.95 1.00 0.86

CIII13-14b2

Thr→ Thr Gly Tyr

0.85 0.85 1.1 0.91

CIII13-14d

Ser→ Ile Asn Asn Ala Ser Leu

0.84 0.93 1.00 1.00 1.05 0.84 0.93

⟵——Ln2——⟶|⟵L2—⟶|

1.03 0.98 1.00 1.00 1.00 0.97 1.01

⟵ Llb ——⟶|⟵Lla—⟶|

0.95 1.02 0.99 0.98 1.02

The CIII15-16 region

The CIII15-16 region was purified by electrophoresis at pH 3.5 for 40 minutes producing CIII15 and CIII16, two closely related peptides, CIII16 being a one amino acid extension of CIII15. CIII16 was a peptide that had been sequenced previously (T12). Below are the compositions and sequences of the two peptides. No N-terminal residue could be determined for CIII15 and only the first residue of CIII16 could be successfully determined by the "Dansyl-Edman" method. No explanation can be given for the failure of the dansyl chloride reagent to

react with some of the N-terminal asparagine residues encountered, a phenomenon frequently observed in this laboratory. There had previously been some doubt about the last three residues in the sequence. To determine the correct order, a peptic digest was done using the same conditions as described earlier. A fragment, CIII15P1, was isolated and the sequence was verified.

CIII15	Asn	Ile	Val	Gly	Gly	Ile	Glu	Tyr
	1.14	0.88	0.86	1.06	1.06	0.88	1.12	0.94
						⟶ P1 ⟵		
						1.02	0.98	0.82

CIII16	Ala	Asn	Ile	Val	Gly	Gly	Ile	Glu	Tyr
	0.92	1.03	0.85	0.77	1.00	1.00	0.85	0.97	0.87

The CIIIn region

This region was separated by pH 1.8 electrophoresis for 70 minutes producing bands CIIIn1, CIIIn2, CIIIn3, CIIIn4, CIIIn5 and CIIIn6. CIIIn2 and CIIIn4 were not isolated in sufficient amounts for further study. CIIIn1 was part of the peptide Tn4 isolated previously and was subjected to the "Dansyl-Edman" method, as were CIIIn3 and CIIIn5. The amino acid analysis of CIIIn3 was not completely satisfactory; however, the N-terminal result was not in contradiction with the suggested sequence. Since this was a neutral peptide, the aspartic acid must exist as an amide. CIIIn6 provided a good overlap between Tn5 and τIVn1d (see Table 4-1). To verify this sequence, a tryptic digest was done by dissolving the peptide in 0.05 M NH_4OH, pH 8.0, and adding a solution of

trypsin in 0.05 M NH_4OH (peptide:enzyme ratio was 100:1). The solution was incubated at 37°C for 5 hours and the resulting peptides purified by electrophoresis at pH 6.5 and pH 1.8. This peptide and its tryptic fragments are shown below along with the other sequences obtained from the CIIIn region.

CIIIn1	Ala→	Glu→	Gly→	Ala→	Val	Arg				
	0.95	1.05	1.11	0.95	1.02	0.93				
CIIIn3	Gly→	Asn	Phe							
	1.00	0.70	1.05							
CIIIn5	Val→	Ser→	Leu→							
	1.10	0.92	0.98							
CIIIn6	Glu→	Arg	Leu	Gln	Pro	Ile	Leu	Ser	Gln	Tyr
	0.95	0.92	1.08	0.95	0.94	0.92	1.08	1.02	0.95	0.81
	⊢→Tn1⊣		⊢→→→			Tn3 ⊣				
	0.96	1.05	0.90	1.00	0.99	1.00	0.93	0.92	1.00	0.93

(c) Fraction IV

Fraction IV was first subjected to electrophoresis at pH 6.5 for 80 minutes producing bands CIV1, CIV2, CIV3, CIV4, CIV5, CIV6, CIVn, CIV7-8, CIV9 and CIV10 upon cadmium-ninhydrin staining.

CIV1

This peptide was further purified by pH 1.8 electrophoresis for 40 minutes and subjected to the "Dansyl-Edman" procedure. It supplied the overlapping sequence between T3-7b1 and T10 (see Table 4-1) isolated earlier.

CIV1 Lys → Gly → Phe

0.88 1.18 1.00

CIV2

CIV2 was purified by pH 3.5 electrophoresis for 50 minutes. This peptide provided an overlap between T3-7a1 and Tn1-3d sequenced earlier. Not only was the "Dansyl-Edman" procedure utilized but due to some doubt about the structure of Tn1-3d, a tryptic digest of CIV2 was carried out under the same conditions as previously outlined. The resulting peptides were purified by electrophoresis at pH 6.5 and pH 1.8 and confirmed the sequence shown below. Although the amino acid analysis of CIV2T1 is not acceptable, the critical portion of CIV2 was the fragment CIV2Tn. From the original analysis of CIV2 it is apparent that only two alanine residues are present.

CIV2 Ala → Ala → Arg → Val → Phe

0.91 0.91 1.00 1.03 1.05

⟵ T1 ⟶ ⟵ Tn ⟶

1.45 1.45 0.87 1.00 1.00

CIV3

CIV3 was isolated after purification by electrophoresis at pH 1.8 for 40 minutes and provided a slight extension of the known peptide Tn5. The sequence of CIV3 was determined by the "Dansyl-Edman" procedure.

CIV3 Arg → Ser → Ser → Leu → Phe

0.93 1.00 1.00 1.05 1.03

CIV4

The CIV4 region was purified by electrophoresis at pH 1.8 for 40 minutes producing two bands upon cadmium-ninhydrin staining, CIV4a and CIV4b. CIV4a was shown to be only a contaminating amino acid. CIV4b was part of a previously elucidated peptide (CDPD2) and its sequence, shown below, was confirmed by the "Dansyl-Edman" method.

CIV4b	Ser →	Gly →	Arg →	Thr	Thr	Gly	Tyr
	1.11	1.01	1.04	0.91	0.91	1.01	0.78

CIV5 and CIV6

CIV5 and CIV6 were both purified in the same way, namely by electrophoresis at pH 1.8 for 40 minutes. The two peptides are closely related, differing only in a terminal tryptophan residue. The sequence of these peptides was known previously since CIV6 is the same peptide as CDPB2 and the tryptophan residue in CIV5 must come at the C-terminal end since the peptide was isolated from a chymotryptic digest. The molar ratio for tryptophan is not included in the composition because of its destruction during hydrolysis. Its presence is detected by Erlich's reagent (freshly prepared 1% p-dimethylaminobenzaldehyde in 90% acetone, 10% conc. HCl).

CIV5	Cys →	Ser	Val	Gly	Phe	Trp
	0.47	0.84	0.92	1.02	0.86	+

CIV6	Cys →	Ser	Val	Gly	Phe
	0.41	0.90	0.87	1.05	0.75

The CIV7-8 region

This band was separated by pH 1.8 electrophoresis for 1 hour. Of the three resulting bands, only CIV7-8a was recovered in adequate amounts for further characterization. Its sequence was determined and is presented below. From its small size and very low mobility at pH 6.5, it is apparent that the glutamic acid residue must be in the amide form.

CIV7-8a	Ser →	Gln →	Ala →
	0.98	1.04	0.96

CIV9

CIV9 was purified by electrophoresis for 40 minutes at pH 1.8 and the sequence determined by the "Dansyl-Edman" procedure. From considerations of its mobility at pH 6.5 and its cadmium-ninhydrin color it is apparent that the peptide contains the aspartic acid in the amide form.

CIV9	Asn →	Gly →	Ser →	Ser →	Phe →
	0.91	1.17	0.97	0.97	0.85

CIV10

This peptide was purified by pH 3.5 electrophoresis for 50 minutes and was found to be the same peptide as T12 isolated previously. Again the "Dansyl-Edman" method failed after the first step.

CIV10	Ala →	Asn	Ile	Val	Gly	Gly	Glu	Ile	Tyr
	0.84	0.99	0.75	0.61	1.00	1.00	1.02	0.75	0.82

The CIVn region

The CIVn region was separated and purified by pH 1.8 electrophoresis for 70 minutes producing CIVn1, CIVn2, CIVn4, CIVn5 and CIVn8. CIVn3, CIVn6 and CIVn7 were not isolated in sufficient amounts for further study, but appeared to be free amino acids. CIVn1 and CIVn2 were sequenced by the "Dansyl-Edman" method. CIVn4 stained for tryptophan using Erlich's reagent but was not isolated in a large enough quantity for a satisfactory amino acid analysis.

CIVn5 was sequenced by the Dansyl-Edman method. From the amino acid analysis it was thought that the peptide had the composition (Ser_2 Gly Val_2 Leu_2 Trp), tryptophan being determined by Erlich's reagent. The "Dansyl-Edman" method worked very well using small amounts of peptide through the first three residues, Val Ser Leu, then failed to give any result in the next step. Repetition of the N-terminal determination using more material produced only a very weak glycine spot, which could easily have been due to glycine contamination. This amount of glycine is not infrequent in N-terminal determinations. The sudden change in behavior forced the tentative conclusion that the peptide is actually a tetrapeptide with the sequence Val Ser Leu Trp, and the glycine in the analysis represents only a high level of glycine contamination. The tetrapeptide status is also more consistent with the electrophoretic mobility at pH 1.8, which is rather high for an octapeptide with no charged residues.

CIVn8 was thought to be a part of a previously sequenced peptide Tn7a and was confirmed as such by the "Dansyl-Edman"

procedure. Since it is a neutral peptide, the aspartic acid must exist as asparagine. The sequences and compositions of the CIVn peptides are shown below.

CIVn1	Gly→	Leu				
	0.92	1.08				
CIVn2	Ser→	Gly				
	1.11	0.89				
CIVn5	Val→	Ser→	Leu→	Trp		also Gly
	0.90	1.00	0.97	+		0.59
CIVn8	Val→	Thr→	Ala→	Asn	Tyr	
	0.90	0.94	1.00	1.04	0.84	

(d) Fraction V

Fraction V was separated by electrophoresis at pH 6.5 for 3 hours. The resulting bands were CV1, CV2, CV3-4 and CV5.

CV1 was further purified by electrophoresis at pH 1.8 for 80 minutes. Since it was suspected that this peptide was an extension of τIVn1a sequenced earlier it was decided to digest it with trypsin (under the same conditions as earlier). The two resulting fragments confirmed the sequence shown below. CV1Tn2 stained for tryptophan using Erlich's reagent.

The CV3-4 region was separated by electrophoresis at pH 1.8 for 50 minutes and CV5 was purified by electrophoresis for 40 minutes at pH 1.8. Both CV5 and CV3 were subjected to the "Dansyl-Edman" procedure but CV2 and CV4 proved to be only free amino acids. The mobility of CV5 dictates that the glutamic acid residue be in the amide form.

Table 3-1

Peptides Isolated from the Chymotryptic

Amino Acid

Peptide	Lys	His	Arg	Asp	Thr	Ser	Glu	Pro	Gly	Ala	AECys[3]	Val
CIII1			1.84			0.91			1.25			
CIII3a			1.04							1.78		
CIII3b	0.97		0.70		1.13				2.32	1.08		
CIII5-6a			0.91						1.10			
CIII5-6b			1.06		0.77				0.74	2.00		
CIII7-8a			1.04			0.77	1.15					
CIII7-8b			1.13		0.82							1.10
CIII9		0.91			0.93				1.00	1.04		0.88
CIII10	1.07			2.02	0.96					1.80		1.01
CIII11				1.10	0.82		0.89		0.89	1.07	0.39	
CIII13-14b1					1.00	1.12			0.95			0.86
CIII13-14b2					1.70				1.10			
CIII13-14d				2.00		1.68				1.05		
CIII15				1.14			1.12		2.12			0.86
CIII16				1.03			0.97		2.00	0.92		0.77
CIIIn1			0.93				1.05		1.11	1.90		1.02
CIIIn3				0.70					1.00			
CIIIn5						0.92						1.10
CIIIn6			0.92			1.02	2.85	0.94				

Digest of S-Aminoethylated α-Lytic Protease

Composition

Met	Ile	Leu	Tyr	Phe	Mobility at pH 6.5	Cadmium Ninhydrin Color	% Yield	Other Comments
					0.85[1]	red	4.0	
					0.70[1]	red	6.6	
				0.84	0.70[1]	red	2.7	
					0.58[1]	yellow	1.1	
					0.58[1]	red	0.3	
					0.53[1]	yellow	11.1	
					0.53[1]	red	1.6	
					0.45[1]	red	19.1	Stains for his
			0.82		0.40[1]	red	5.3	Stains for tyr
0.63					0.35	red	2.1	
		0.97			0.09[2]	yellow-orange	3.1	
			0.91		0.09[2]	yellow	2.7	Stains for tyr
	0.93	0.93			0.09[2]	orange	20.0	
	1.76		0.94		0.25[2]	red	2.7	Stains for tyr
	1.70		0.87		0.25[2]	red	10.7	Stains for tyr
					0.00	orange-red	8.5	
				1.05	0.00	orange	2.2	
		0.98			0.00	red	9.8	
	0.92	2.16	0.81		0.00	red	16.0	Stains for tyr

(continued...)

Table 3-1

Peptide	Lys	His	Arg	Asp	Thr	Ser	Glu	Pro	Gly	Ala	AECys[3]	Val
CIV1	0.88								1.18			
CIV2			1.00							1.82		1.03
CIV3			0.93			2.00						
CIV4b			1.04		1.82	1.11			2.02			
CIV5						0.84			1.02		0.47	0.92
CIV6						0.90			1.05		0.41	0.87
CIV7-8a						0.98	1.04			0.96		
CIV9				0.91		1.94			1.17			
CIV10				0.99			1.02		2.00	0.84		0.61
CIVn1									0.92			
CIVn2						1.11			0.89			
CIVn5						1.00						0.90
CIVn8				1.04	0.94					1.00		0.90
CV1			0.87	1.92				0.94	1.02	1.00		
CV3									0.84			
CV5						1.07	0.92					

[1] calculated with respect to lysine

[2] calculated with respect to aspartic acid

[3] S-β-aminoethylcysteine

(continued)

Met	Ile	Leu	Tyr	Phe	Mobility at pH 6.5	Cadmium Ninhydrin Color	% Yield	Other Comments
				1.00	0.59[1]	red	1.2	
				1.05	0.50[1]	orange-red	12.0	
		1.05		1.03	0.41[1]	red	6.7	
			0.78		0.35[1]	yellow	6.0	Stains for tyr
				0.86	0.31[1]	red	4.9	Stains for trp
				0.75	0.25[1]	orange-red	12.0	
					0.08[2]	yellow	4.7	
				0.85	0.13[2]	yellow	3.0	
	1.50		0.82		0.28[2]	red	4.0	Stains for tyr
		1.08			0.00	red	6.7	
					0.00	red	4.7	
	0.97				0.00	red	4.7	High gly impurity Stains for trp
			0.84		0.00	red	1.9	Stains for tyr
					0.05[1]	yellow-orange	22.2	Stains for trp
	1.09				0.00	yellow	0.5	
			0.98		0.07[2]	yellow	2.7	Stains for tyr

CVI

Pro	Gly	Asn	Asp	Arg	Ala	Trp
0.94	1.02	0.86	0.86	0.87	1.00	+
		Tn1			Tn2	
0.96	1.00	0.98	0.98	1.00	1.00	+

CV3

Gly	Leu
0.84	1.09

CV5

Ser	Gln	Tyr
1.07	0.92	0.98

5. Discussion

It can be appreciated from the results of this chapter that the experimental methods used had both advantages and disadvantages. One of the disadvantages first became apparent in the conversion of the protein to the aminoethylated derivative, which proved to be a very inconsistent procedure in our hands. Much time and effort were put into attempts to utilize several known variations of procedures for aminoethylation of proteins and some modifications of these procedures were attempted. The recovery of S-β-aminoethylcysteine ranged from a minimum of 1.25 residues to a maximum of 5.6 residues (theoretical = 6) over the range of procedures. However, even a single procedure did not consistently give the same yield. The same situation has been encountered by other workers in this and other laboratories (58, 62), and no explanation can be given at this time for the lack of reproducibility of any single aminoethylation experiment. The yield of S-β-aminoethylcysteine (3.8 residues) obtained in the protein used for digestion in this work represented

an average result for this laboratory.

Other disadvantages of the method used were in the use of chymotrypsin as the digesting enzyme and the conditions under which it was used. Although this protease produced some of the desired overlaps, the imperfect specificity caused some problems. A large number of peptides were isolated from the digest, a result of many partial splits and hydrolysis at residues where chymotrypsin is not as efficient as it is at aromatic sites. Partial splitting of this nature results in poor yields of many peptides, a condition which was observed in the present study. The problem of impure peptides could also be at least in part a result of partial splits. As mentioned earlier, some of the peptides obtained could not be satisfactorily purified. This could have been due to the fact that each unpurifyable band might have been two peptides closely related in composition which resulted from a partial split of a residue terminal to the sequence. The new fragment thus formed would have one less amino acid than the parent peptide and if it were large, its mobility would not be appreciably affected. Thus the two peptides could have the same electrophoretic mobility at almost any pH.

Possibly the chymotryptic hydrolysis could be modified to eliminate or at least minimize some of its disadvantages. The first obvious change would be to use less rigorous conditions for digestion. The enzyme:protein ratio could be lowered to 1:100 and the hydrolysis time shortened to perhaps two hours rather than digesting for five hours as was done in the present study. This should greatly lower the number of

peptides obtained. Since several peptides with a C-terminal arginine residue were isolated from the digest, it is likely that the chymotrypsin used was contaminated with trypsin. The extra splitting due to tryptic hydrolysis could be largely eliminated by treatment of the chymotrypsin with the trypsin inhibitor TLCK.

The major advantage of the method was simply that it provided many of the desired results. The S-aminoethylated derivative of α-lytic protease proved to be soluble throughout its preparation and digestion. The digest yielded a number of valuable overlapping peptides for the tryptic fragments previously elucidated and thus permitted the alignment of these peptides to further extend the known amino acid sequence of the protein. Although the complete sequence of α-lytic protease cannot be determined from the data thus far collected, significant portions of the molecule can be pieced together. This is shown more clearly in Table 4-1.

CHAPTER IV

CONCLUSIONS

1. Evolution of the serine proteases

Postulating evolution of two proteins from a common ancestor can be done reliably only when the complete amino acid sequences are known; that is to say analogous proteins (proteins with similarities in function but not structure) are no indication of common ancestry but homologous proteins (those which possess similarities in amino acid sequence) do suggest a common ancestral gene. Homologies of certain short sequences are sometimes used as indications of homologous proteins but such comparisons must be cautiously interpreted. Amino acid composition cannot be a reliable criterion for homology although it has been shown that homologous proteins do in fact possess similar amino acid compositions (38).

With the sequence data available, a crude hypothetical evolutionary tree showing the successive gene duplications which led to the present structures of many of the homologous serine proteases has been constructed (39) and is shown in Figure 4-1. It can be assumed that since the generation time for bacteria is much shorter than for higher animals, the subtilisin group evolved most recently, an assumption which may be supported by the relatively high degree of homology in these species of proteins.

The ancestral gene for the DFP inhibited esterases with the active centre sequence -Glu Ser* Ala- and the gene for the serine proteases having an active centre sequence -Asp Ser* Gly- presumably coded for a primitive esterase having

Figure 4-1

Hypothetical Evolutionary Descent of the
Serine Proteases and Esterases

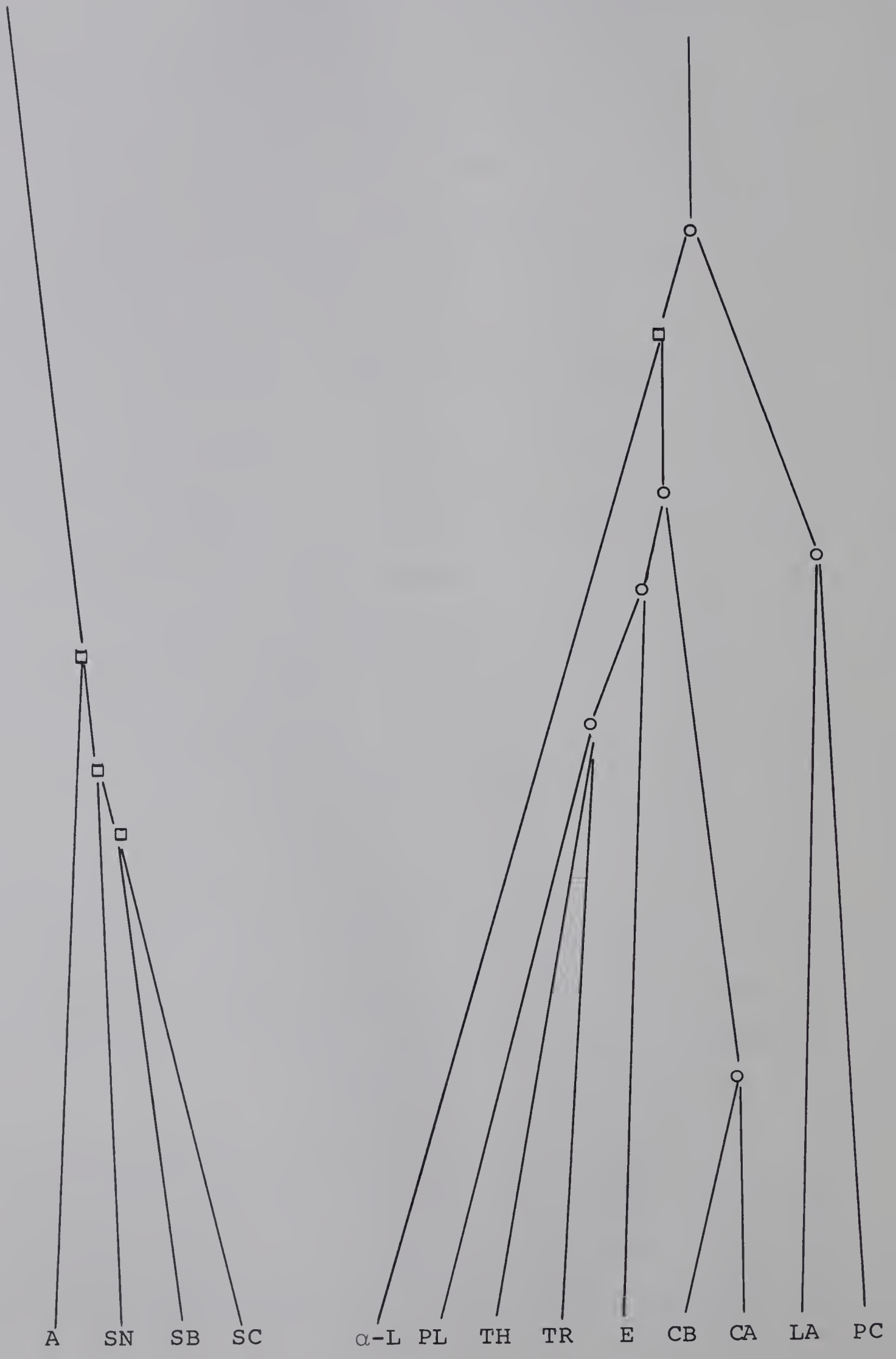


Figure 4-1 Legend

□	separation of genes due to speciation
O	gene duplication

Enzyme code

A	aspergillus protease
SN	subtilisin novo
SB	subtilisin BPN' (nagarse)
SC	subtilisin Carlsberg
TR	trypsinogen
TH	prothrombin
PL	plasminogen
E	pro-elastase
CB	chymotrypsinogen B
CA	chymotrypsinogen A
α-L	α-lytic protease of Sorangium
LA	liver ali-esterase
PC	pseudocholinesterase

an active centre serine and an active sequence of one of the four combinations of $\begin{matrix}\text{Glu}\\\text{Asp}\end{matrix}$ Ser* $\begin{matrix}\text{Ala}\\\text{Gly}\end{matrix}$ (40). This gene could then have undergone duplication giving rise to the two genes which led to the parent genes of the DFP inhibited esterases and the serine proteases. Later, but after the active sequence -Asp Ser* Gly- had been established in the serine protease line, the soil bacterium Sorangium sp. separated from the main evolutionary line toward higher animals, and thus the gene for α-lytic protease began to evolve independently from the rest of the serine proteases. The next major events could have been the closely spaced gene duplications resulting in the formation of two new genes which were to give rise to the chymotrypsins and pro-elastase.

An argument favoring a parent trypsin-like enzyme as the earliest serine protease is the one suggested by the fact that if the earliest enzyme resembled the modern proteases, it must have been able to activate its own zymogen, and trypsin is the only enzyme in this series that has this property. Possible evidence against this has been presented by Jukes (41) who has suggested that, judging by its longer length, the gene for chymotrypsinogen could be older. The most recent gene duplication known in the serine protease line was the one giving rise to the gene for chymotrypsinogen B. This may have occurred as recently as 400×10^6 years ago although it is difficult to assign a time scale due to a lack of information about the rate of evolution of this group of proteins. The number of amino acid differences in homologous proteins from two species has been found to be roughly proportional to the time since divergence of the

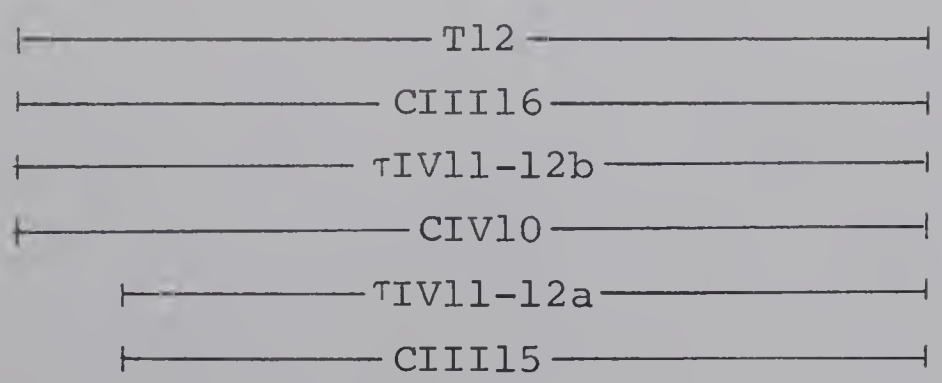
species. However, Hill and Buettner-Janusch (42) emphasize that the rate of substitution depends to a large extent on the proteins themselves. Another important factor would be the generation time of the species from which the proteins were obtained.

2. The structure of α-lytic protease

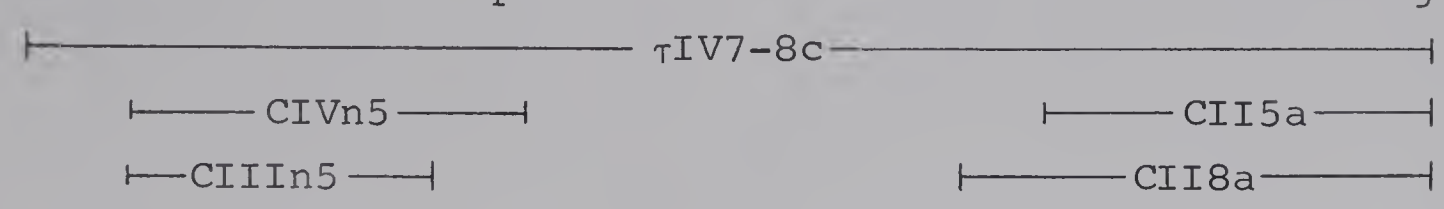
Table 4-1 presents all the postulated unique sequences of α-lytic protease to the present time. The varied nomenclature is the result of the many different studies done on the protein. All peptides beginning with "T" are the tryptic peptides of the reduced and carboxymethylated protein described in Chapter II of this thesis. "CII","CIII", "CIV" and "CV" peptides are the Sephadex separated fractions of the chymotryptic digest of the reduced and aminoethylated enzyme described in Chapter III. All "τ" peptides are the fragments arising from the trypsin digest of the S-aminoethylated α-lytic protease performed in this laboratory by Dr. N. Nagabhushan. The "CDP" peptides are cysteic acid peptides (see Table 2-1) isolated previously from a peptic digest and the diagonal procedure of Brown and Hartley (56). "CNBr" peptides were isolated by Whitaker (64) using cyanogen bromide cleavage of native α-lytic protease. "CNBr-A" and "CNBr-B" resulted from Sephadex separation of the cyanogen bromide treated material. CNBr-A was reduced, carboxymethylated and separated by column chromatography on Sephadex producing CNBr-A1, CNBr-A2, CNBr-A3 and CNBr-A4. One of the major fractions, CNBr-A4, was digested with trypsin producing CNBr-A4T peptides. All sequence work on CNBr, τI

Table 4-1 Summary of Postulated Unique Sequences of α-Lytic Protease

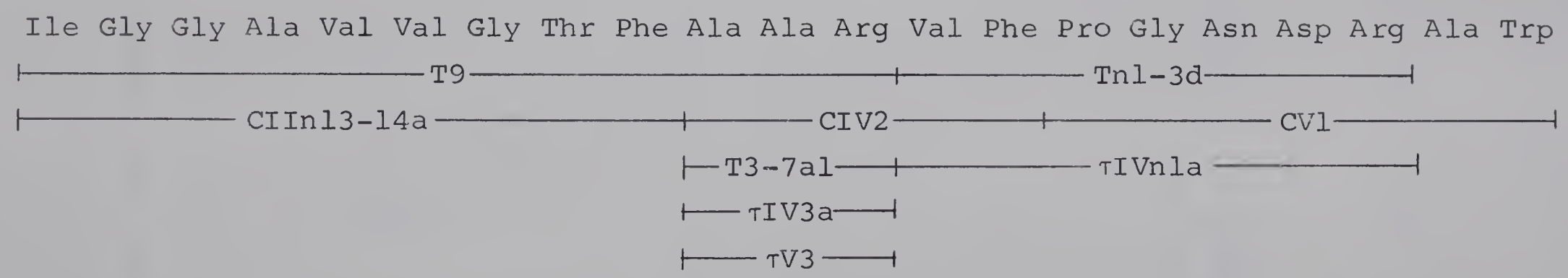
1. Ala Asn Ile Val Gly Gly Ile Glu Tyr



2. Ala Val Ser Leu Trp Thr Ser Ala Gln Thr Leu Leu Pro Arg



3. Ile Gly Gly Ala Val Val Gly Thr Phe Ala Ala Arg Val Phe Pro Gly Asn Asp Arg Ala Trp



4. Arg Gly Ala Thr Lys Gly Phe Val Thr Ala Gly His Cys Gly Thr Val Asn Ala Thr Ala Arg

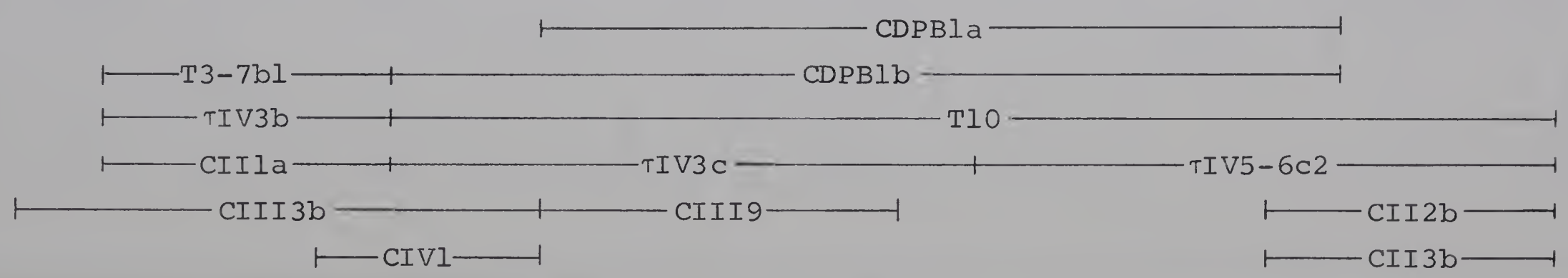
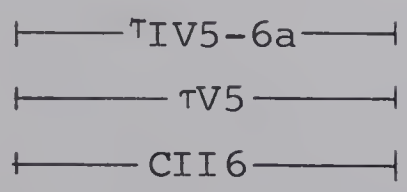
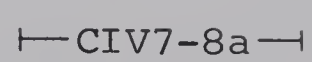


Table 4-1 (continued)

5. Ser Val Thr Arg



6. Ser Gln Ala



7.

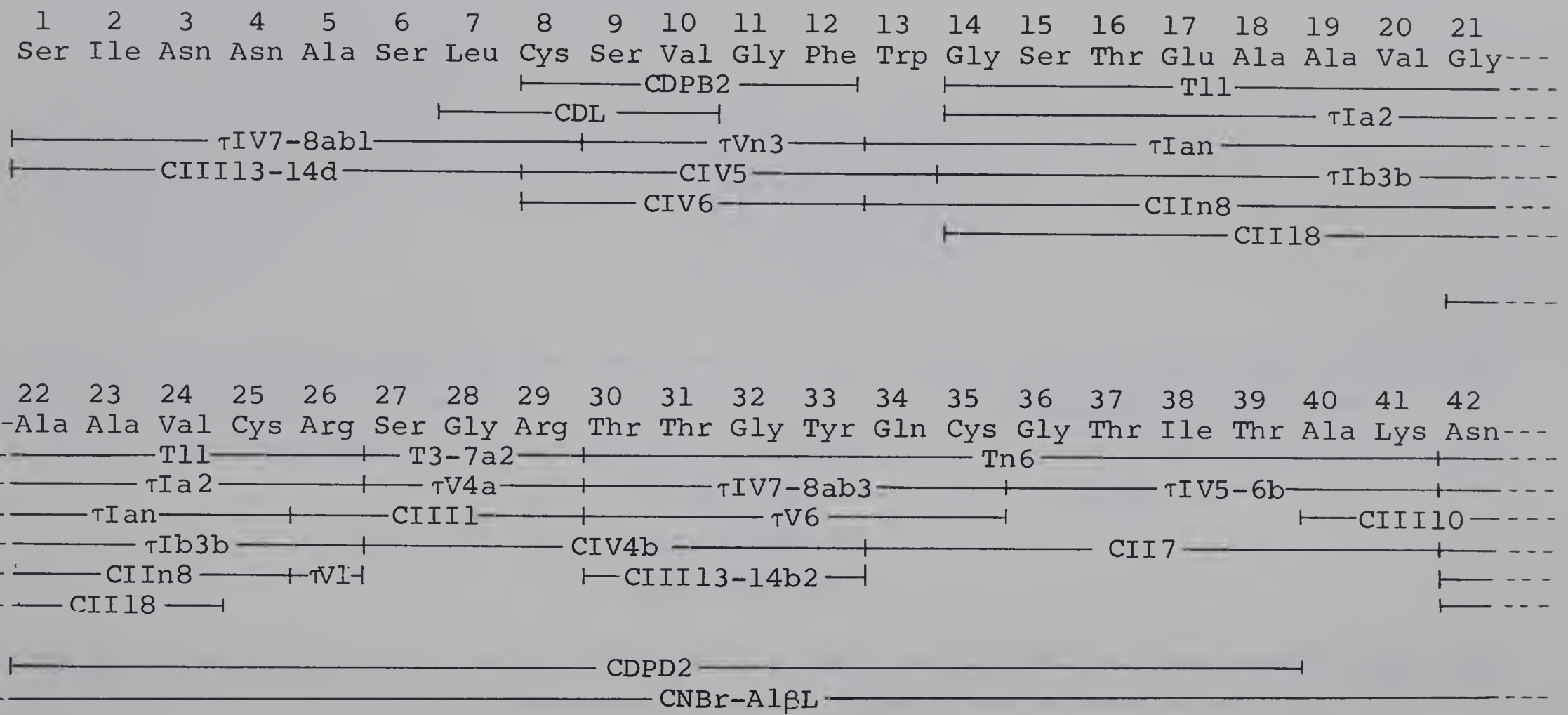


Table 4-1 (continued)

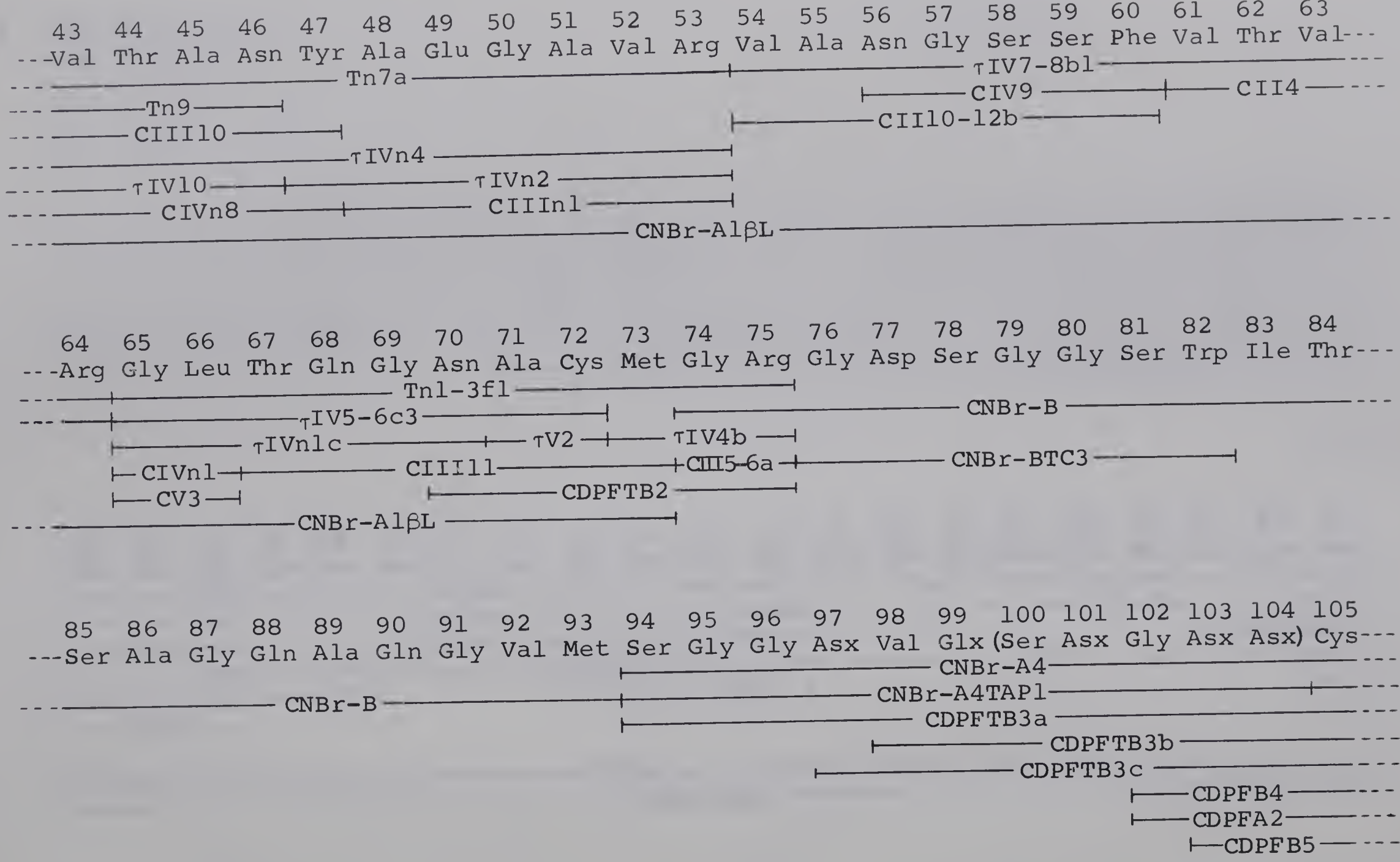
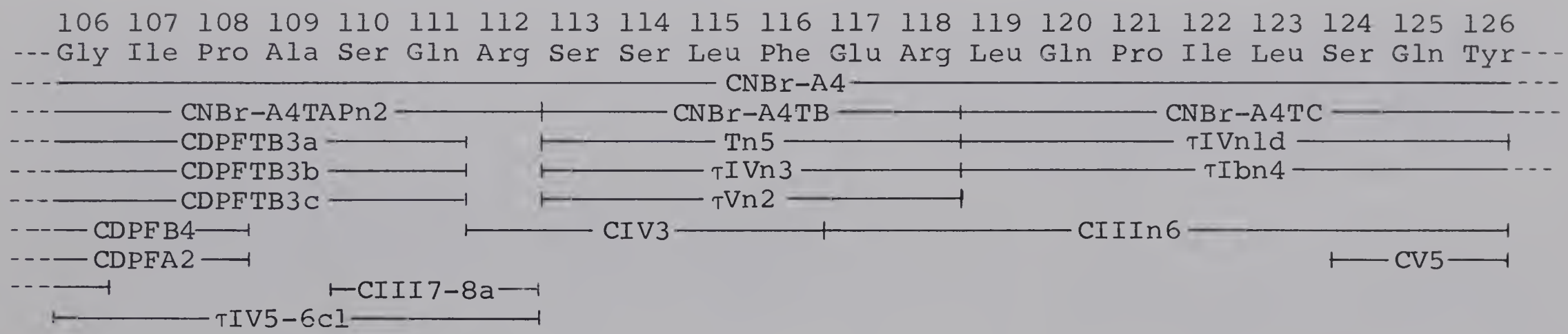
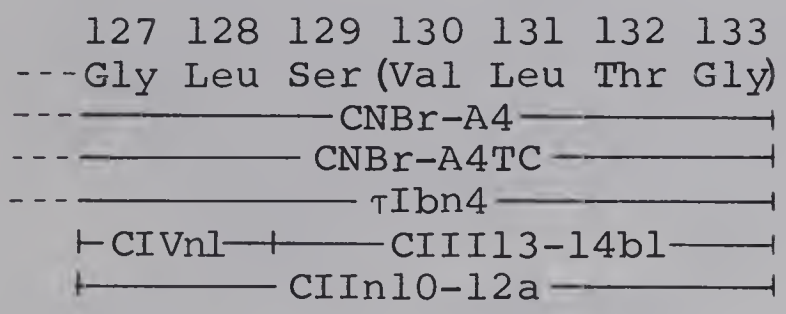


Table 4-1 (continued)





and CII peptides was done in this laboratory by Dr. M. Olson (57).

It should be emphasized that the sequences tabulated are postulated sequences. The evidence for the order of some fragments, especially in peptide 7 in Table 4-1 is not complete. Since α-lytic protease has an N-terminal alanine residue, peptide 1 or peptide 2 is likely the N-terminal portion of the molecule. Peptide 7 is of particular interest and for convenience is numbered from its N-terminal end. CNBr-A4 proved to be the C-terminus of the protein. This fragment was the only peptide which after cyanogen bromide cleavage, reduction and carboxymethylation, showed no homoserine to be present upon analysis. Since cyanogen bromide cleavage of the native enzyme (without reduction and carboxymethylation) released the peptide CNBr-B, and the rest of the molecule as a second fragment, the cystine residue numbered 72 is linked in a disulfide bridge with cystine-105. This agrees with the previous determination by the diagonal procedure. Of additional interest in peptide 7 is the postulation that the "active serine" sequence is contained within it.

3. Comparison of the structures of chymotrypsin and α-lytic protease

It was indicated previously that X-ray studies have demonstrated the near total exclusion of polar residues from the interiors of protein molecules. When the amino acid sequences of some 18 globin chains from various species of myoglobin and hemoglobin were compared, this feature was expressed

in a pattern of 30 sites where only non-polar residues occurred (63). A considerable variety of replacements was permissible at these sites as long as the non-polar character was maintained. Similar comparisons on the chymotrypsinogens A and B and trypsinogen by Smillie et al. (49) demonstrated that such a pattern also existed for these enzymes. Hence when comparing the structure of α-lytic protease with that of chymotrypsin, attention should be paid to the pattern of invariant hydrophobic residues. A similarity in the patterns of these molecules could indicate a resemblance in three dimensional shape.

Already mentioned in the introduction of this thesis is the fact that the histidine sequence of α-lytic protease has some homology to the corresponding portion of chymotrypsin. As can be seen from Table 4-2, not only is there a homologous sequence of amino acids around the histidine (the only exception being a conservative replacement in α-lytic protease of glycine for alanine in the 56 position) but the pattern of invariable non-polar residues in this area of trypsinogen and the chymotrypsinogens is almost perfectly adhered to in the α-lytic protease molecule. A further comparison of this area cannot be made due to lack of knowledge about the sequence of α-lytic protease on either side of the area shown in Table 4-2.

It has been possible at this time to tentatively align a large portion (133 residues) of the polypeptide chain representing some two-thirds of the C-terminal part of the molecule. When this large fragment is compared with the trypsin, chymotrypsin A and chymotrypsin B molecules by aligning the "active serine" sequences, the disulfide bridge and the C-terminal

Table 4-2

The Histidine Sequences of Some Proteolytic Enzymes Including a Comparison of Invariant Non-Polar Residues in the Mammalian Serine Proteases with Those of α-Lytic Protease

The following abbreviations are used: T for trypsinogen, CA for chymotrypsinogen A, CB for chymotrypsinogen B, and αLP for α-lytic protease.

The invariant hydrophobic residues are shown in the regions enclosed by solid lines.

The numbering system is that used in Table 1-2.

	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66
CA	Leu	Ile	Asn	Glu	Asn	Trp	Val	Val	Thr	Ala	Ala	His	Cys	Gly	Val	Thr	Thr	Ser	Asp	Val	Val
CB	Leu	Ile	Ser	Glu	Asp	Trp	Val	Val	Thr	Ala	Ala	His	Cys	Gly	Val	Thr	Thr	Ser	Asp	Val	Val
T	Leu	Ile	Asn	Ser	Gln	Trp	Val	Val	Ser	Ala	Ala	His	Cys	Tyr	Lys	Ser	Gly	Ile	Gln	Val	Arg
αLP	Arg	Gly	Ala	Thr	Lys	Gly	Phe	Val	Thr	Ala	Gly	His	Cys	Gly	Thr	Val	Asn	Ala	Thr	Ala	Arg

sequence as in Table 4-3, some similarities are again obvious. (The "active" serine is residue number 198 in this numbering scheme.) The pattern of invariant hydrophobic residues of the mammalian serine proteases in this area is again approximated by the bacterial enzyme. Since α-lytic protease is a smaller molecule than the mammalian proteins, it is to be expected that a greater proportion of the residues will be on the exterior of the molecule. This is indeed the case when it is assumed that all polar residues extend towards the exterior of the molecule.

Another comparison between these molecules proved to be of interest. The number of amino acid residues found between the two cystines of the "disulfide loop" containing the active site is nearly the same in the two enzymes. If the distance in amino acid residues between the "active" serine and either cystine residue in that "disulfide loop" is compared, it can be seen that this distance is two residues greater (to either cystine from the serine) in α-lytic protease than in chymotrypsin. Although these intervening residues are not identical in the two proteins, the sizes of the "disulfide loops" should be almost the same in both molecules and in both cases are about an equal distance from the C-terminus. The "active" serine residues themselves are in nearly identical positions, the serine of α-lytic protease being 55 residues from the C-terminus and the corresponding residue of chymotrypsin being 51 residues from its C-terminal end. These similarities are undoubtedly of importance in the stereochemistry of the catalytic reaction and in the three-dimensional structures of these

Table 4-3

A Comparison of the Invariant Non-Polar Residues at the C-Terminal End of Some Mammalian Serine Proteases with α-Lytic Protease

The following abbreviations are used: T for trypsinogen, CA for chymotrypsinogen A, CB for chymotrypsinogen B and αLP for α-lytic protease.

The invariant hydrophobic residues are shown in the regions enclosed by solid lines.

The numbering system is that used in Table 1-2.

	127	128	129	130	313	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147
CA	Ala	Ser	Asp	Asp	Phe	Ala	Ala	Gly	Thr	Thr	Cys	Val	Thr	Thr	Gly	Trp	Gly	Leu	Thr	Arg	Tyr
CB	Ala	Asp	Glu	Asp	Phe	Pro	Ala	Gly	Met	Leu	Cys	Ala	Thr	Thr	Gly	Trp	Gly	Lys	Thr	Lys	Tyr
T	-	Ser	Cys	Ala	Ser	-	Ala	Gly	Thr	Gln	Cys	Leu	Ile	Ser	Gly	Trp	Gly	Asn	Thr	Lys	Ser
αLP	Ser	Ile	Asn	Asn	Ala	Ser	Leu	Cys	Ser	Val	Gly	Phe	Trp	Gly	Ser	Thr	Glu	Ala	Ala	Val	Gly

	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168
CA	Thr	Asn	Ala	Asn	Thr	Pro	Asp	Arg	Leu	Gln	Gln	Ala	Ser	Leu	Pro	Leu	Leu	Ser	Asn	Thr	Asn
CB	Asn	Ala	Leu	Lys	Thr	Pro	Asp	Lys	Leu	Gln	Gln	Ala	Thr	Leu	Pro	Ile	Val	Ser	Asn	Thr	Asp
T	Ser	Gly	Thr	Ser	Tyr	Pro	Asp	Val	Leu	Lys	Cys	Leu	Lys	Ala	Pro	Ile	Leu	Ser	Asn	Ser	Ser
αLP	Ala	Ala	Val	Cys	Arg	Thr	Thr	Gly	Tyr	Gln	Cys	Thr	Ile	Thr	Ala	Lys	Asn	Val	Thr	Ala	Asn
												Gly									

	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189
CA	Cys	Lys	Lys	Tyr	Trp	Gly	Thr	Lys	Ile	Lys	Asp	Ala	Met	Ile	Cys	Ala	Gly	Ala	-	Ser	Gly
CB	Cys	Arg	Lys	Tyr	Trp	Gly	Ser	Arg	Val	Thr	Asp	Val	Met	Ile	Cys	Ala	Gly	Ala	-	Ser	Gly
T	Cys	Lys	Ser	Ala	Tyr	Pro	Gly	Gln	Ile	Thr	Ser	Asn	Met	Phe	Cys	Ala	Gly	Tyr	Leu	Glu	Gly
αLP	Tyr	Ala	Glu	Gly	Ala	Val	Arg	Val	Ala	Asn	Gly	Ser	Ser	Phe	Val	Thr	Val	Arg	Gly	Leu	Thr

	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210
CA	Val	-	Ser	Ser	Cys	Met	Gly	Asp	Ser	Gly	Gly	Pro	Leu	Val	Cys	Lys	Lys	Asn	Gly	Ala	Trp
CB	Val	-	Ser	Ser	Cys	Met	Gly	Asp	Ser	Gly	Gly	Pro	Leu	Val	Cys	Gln	Lys	Asn	Gly	Ala	Trp
T	Gly	Lys	Asn	Ser	Cys	Gln	Gly	Asp	Ser	Gly	Gly	Pro	Val	Val	Cys	Ser	Gly	Lys	-	-	-
αLP	Gln	Gly	Asn	Ala	Cys	Met	Gly	Asp	Ser	Gly	Gly	Ser	Trp	Ile	Thr	Ser	Ala	Gly	Gln	Ala	Gln
						Gly	Arg														

Table 4-3 (continued)

	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230
CA	Thr	Leu	Val	Gly	Ile	Val	Ser	Trp	Gly	Ser	Ser	Thr	Cys	Ser	Thr	Ser	Thr	-	Pro	Gly
CB	Thr	Leu	Ala	Gly	Ile	Val	Ser	Trp	Gly	Ser	Ser	Thr	Cys	Ser	Thr	Ser	Thr	-	Pro	Ala
T	-	Leu	Gln	Gly	Ile	Val	Ser	Trp	Gly	Ser	-	Gly	Cys	Ala	Gln	Lys	Asn	Lys	Pro	Gly
αLP	Gly	Val	Met	Ser	Gly	Gly	Asx	Val	Glx	(Ser	Asx	Gly)	Cys	Gly	Ile	Pro	Ala	Arg	Ser	Ser
												Asx	Asx				Ser	Gln		

	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249
CA	Val	Tyr	Ala	Arg	Val	Thr	Ala	Leu	Val	Asn	Trp	Val	Gln	Gln	Thr	Leu	Ala	Ala	Asn
CB	Val	Tyr	Ala	Arg	Val	Thr	Ala	Leu	Met	Pro	Trp	Val	Gln	Glu	Thr	Leu	Ala	Ala	Asn
T	Val	Tyr	Thr	Lys	Val	Cys	Asn	Tyr	Val	Ser	Trp	Ile	Lys	Gln	Thr	Ile	Ala	Ser	Asn
αLP	Leu	Phe	Glu	Arg	Leu	Gln	Pro	Ile	Leu	Ser	Gln	Tyr	Gly	Leu	Ser	Val	Leu	Thr	Gly

proteins.

Some problems arise when attempts are made to compare the α-lytic protease sequence with that of other enzymes. This becomes obvious in Table 4-3. The chymotrypsins consist of 245 amino acid residues while α-lytic protease is reported to have only 198 residues (55). Since the complete amino acid sequence of α-lytic protease is not yet known, it is difficult to estimate where the extra 47 residues of the mammalian enzymes could be put into its chain for comparative purposes. Note that in the comparisons presented in Table 4-2 and Table 4-3 small segments of the α-lytic protease chain were deleted for the purpose of maximizing homologies. These small segments are shown below the sequences.

In light of what is known about the α-lytic protease molecule at the present time, it seems quite possible that this bacterial protein has evolved from the same precursor as the mammalian enzymes. Another possibility, of course, is that this similarity in structure could have arisen as a result of convergent evolution. The knowledge of α-lytic protease to date is just sufficient to suggest that it has some likeness to the mammalian serine proteases. Except for the sizes of some of the "disulfide loops", little is known about the three-dimensional structure of the bacterial enzyme. Whitaker has recently shown that optical rotatory dispersion data present no evidence of α-helices in the α-lytic protease molecule (65). The same phenomenon has been found in chymotrypsin; only eight residues at the C-terminal end are coiled. However, a complete comparison of the structures of chymotrypsin and

α-lytic protease will require both total amino acid sequences and a knowledge of their three-dimensional structures, which can hopefully be obtained from X-ray crystallographic procedures. Such a comparison awaits further study.

BIBLIOGRAPHY

1. D.R. Whitaker, Can. J. Biochem., 43, 1935 (1965).
2. L.B. Smillie and D.R. Whitaker, J. Am. Chem. Soc., 89, 3350 (1967).
3. M. Florkin and E.H. Stotz, Editors, Comprehensive Biochemistry, Vol. 13, Elsevier Publishing Co., Amsterdam (1964).
4. B.S. Hartley, Ann. Rev. Biochem., 29, 45 (1960).
5. K.F. Tipton, Biochim. Biophys. Acta, 110, 414 (1965).
6. L.W. Cunningham, Science, 125, 1145 (1957).
7. B. Zerner and M.L. Bender, J. Am. Chem. Soc., 86, 3669 (1964).
8. B. Zerner, R.P.M. Bond and M.L. Bender, J. Am. Chem. Soc., 86, 3674 (1964).
9. M.L. Bender, G.E. Clement, F.J. Kézdy and H.D'A. Heck, J. Am. Chem. Soc., 86, 3680 (1964).
10. F.J. Kézdy, G.E. Clement and M.L. Bender, J. Am. Chem. Soc., 86, 3690 (1964).
11. M.L. Bender, G.E. Clement, C.R. Gunter and F.J. Kézdy, J. Am. Chem. Soc., 86, 3697 (1964).
12. M.L. Bender and F.J. Kézdy, J. Am. Chem. Soc., 86, 3704 (1964).
13. M.L. Bender, F.J. Kézdy and C.R. Gunter, J. Am. Chem. Soc. 86, 3714 (1964).
14. I.B. Wilson, F. Bergmann and D. Nachmansohn, J. Biol. Chem., 186, 781 (1950).
15. M.L. Bender, J. Am. Chem. Soc., 84, 2582 (1962).
16. M.L. Bender and F.J. Kézdy, Ann. Rev. Biochem., 34, 49 (1965).

17. B.W. Matthews, P.B. Sigler, R. Henderson and D.M. Blow, Nature, 214, 652 (1967).

18. B.S. Hartley, J.R. Brown, D.L. Kauffman and L.B. Smillie, Nature, 207, 1157 (1965).

19. B.S. Hartley, Nature, 201, 1284 (1964).

20. A. Furka, L.B. Smillie, K.J. Stevenson and C.O. Parkes, Federation Proc., 25, 789 (1966).

21. K.A. Walsh, K.S.V. Sampath Kumar and D.L. Kauffman, Abstracts, 6th International Congress of Biochemistry, Vol. 2, p.186 (1964).

22. N.K. Schaffer, S.C. May and W.H. Summerson, J. Biol. Chem., 202, 67 (1953).

23. F. Turba and G. Gundlach, Biochem. Z., 327, 186 (1955).

24. K.J. Stevenson, Ph.D. Thesis, University of Alberta (1966).

25. K.A. Walsh and H. Neurath, Proc. Nat. Acad. Sci., 52, 884 (1964).

26. L.B. Smillie and D.R. Whitaker, Unpublished results.

27. J.A. Gladner and K. Laki, J. Am. Chem. Soc., 80, 1263 (1958).

28. B.S. Hartley, M.A. Naughton and F. Sanger, Biochim. Biophys. Acta, 34, 243 (1959).

29. D.R. Whitaker and C. Roy, Can. J. Biochem., 45, 911 (1967).

30. H.S. Jansz, C.H. Posthumus and J.A. Cohen, Biochim. Biophys. Acta, 33, 396 (1959).

31. H.S. Jansz, D. Brons and M.G.P.J. Warringa, Biochim. Biophys. Acta, 34, 573 (1959).

32. F. Sanger, Proc. Chem. Soc., 76 (1963).

33. F. Sanger and D.C. Shaw, Nature, 187, 872 (1960).

34. G. Schoellmann and E. Shaw, Biochem., 2, 252 (1963).

35. E. Shaw, M. Mares-Guia and W. Cohen, Biochem., 4, 2219 (1965).

36. H. Kaplan and D.R. Whitaker, J. Am. Chem. Soc., 89, 3352 (1967).

37. E.L. Smith, F.S. Markland, C.B. Kasper, R.J. DeLange, M. Landon and W.H. Evans, J. Biol. Chem., 241, 5974 (1966).

38. H. Neurath, K.A. Walsh and W.P. Winter, Science, 158, 1638 (1967).

39. D. Gibson, Ph.D. Thesis, University of British Columbia (1967).

40. G.H. Dixon in Essays in Biochemistry, Vol. 2, Academic Press, New York, N.Y., p.147 (1966).

41. T.A. Jukes, Molecules and Evolution, Columbia Univ. Press, New York, N.Y., p.58 (1966).

42. J. Buettner-Janusch and R.L. Hill, Science, 147, 836 (1965).

43. L.B. Smillie and B.S. Hartley, Abstr. 1st Meeting Fed. European Biochem. Soc., Academic Press, London, A30 (1964).

44. S. Wählby, Biochim. Biophys. Acta, 151, 394 (1968).

45. S. Wählby and L. Engström, Biochim. Biophys. Acta, 151, 402 (1968).

46. S. Wählby, Biochim. Biophys. Acta, 151, 409 (1968).

47. W.E. Marshall, R. Manion and J. Porath, Biochim. Biophys. Acta, 151, 414 (1968).

48. M. Dixon and E.C. Webb, The Enzymes, 2nd Ed., Academic Press Inc., New York, N.Y., p.672 (1964).

49. L.B. Smillie, A. Furka, N. Nagabhushan, K.J. Stevenson, and C.O. Parkes, Nature, 218, 343 (1968).

50. P.B. Sigler and D.M. Blow, J. Mol. Biol. (in press).

51. D.R. Whitaker, F.D. Cook and D.C. Gillespie, Can. J. Biochem., 43, 1927 (1965).

52. L. Jurášek and D.R. Whitaker, Can. J. Biochem., 43, 1955 (1965).

53. D.R. Whitaker, C. Roy, C.S. Tsai and L. Jurášek, Can. J. Biochem., 43, 1961 (1965).

54. C.S. Tsai, D.R. Whitaker, L. Jurášek and D.C. Gillespie, Can. J. Biochem., 43, 1971 (1965).

55. L. Jurášek and D.R. Whitaker, Can. J. Biochem., 45, 917 (1967).

56. J.R. Brown and B.S. Hartley, Biochem. J., 101, 214 (1966).

57. M.O.J. Olson, Unpublished results.

58. N. Nagabhushan, Unpublished results.

59. D.G. Smyth, W.H. Stein and S. Moore, J. Biol. Chem., 237, 1845 (1962).

60. M.A. Raftery and R.D. Cole, Biochem. Biophys. Res. Commun., 10, 467 (1963).

61. M.A. Raftery and R.D. Cole, J. Biol. Chem., 241, 3457 (1966).

62. P.E. Wilcox, Private communication.

63. M.F. Perutz, J.C. Kendrew and H.C. Watson, J. Mol. Biol., 13, 669 (1965).

64. D.R. Whitaker, Unpublished results.

65. G.M. Paterson, D.R. Whitaker and P. Morand, Can. J. Biochem., (in press).

66. W.R. Gray in Methods in Enzymology, Vol. XI, Academic Press, New York, N.Y., p. 469 (1967).